Vol. 43, No. 4 July, 1936

PSYCHOLOGICAL REVIEW PUBLICATIONS

Psychological Review

CONTENTS

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Vol. 43, No. 4 July, 1936

The Psychological Review

SUMMATION OF STIMULI AND THE NEURAL CHANGES IN LEARNING

BY M. N. CHAPPELL AND F. H. PIKE
Columbia University

Attempts to understand learning in terms of classical neurophysiology have proven so sterile as to lead many psychologists to conclude that neurophysiology has nothing to offer psychology. Indeed, some have gone so far as to postulate that neurophysiology must eventually be given in terms of psychological processes. The reflex arc, synaptic resistance, canalization and blockage have, to use Sherrington's phrase, proven to be convenient, if mythical concepts. Time was when learning consisted of compounding three neurone arcs and synaptic resistance would explain everything from random activity to intelligence. As the data of psychology have accumulated these neurological monsters have been relegated to their proper realm. But it is also true that some significant data have been accumulated in neurophysiology since the three-neurone arc was postulated. It may be profitable to examine the problem of learning in terms of these data. Two lines of development in neurophysiology seem to have particular significance to the problem of learning. These are, (*a*) the changes in the functional capacity of a conduction system, and (*b*) summation of stimuli. It is the aim of this paper to show that these data will lead to a better conception of learning in many of its aspects.

I

Changes in the Functional Capacity of a Conduction Pathway

Changes in the functional capacity of a conduction pathway are indicated by a multitude of names which, after careful

classification, may be reduced to three—growth, compensation for injury to the nervous system and learning. From long experience we have learned that a multitude of names, or even three, may not necessitate more than a single underlying physiological process for their explanation.

The aim of this section is to show that the change which underlies the acquisition of new forms of reaction, or increase in the magnitude of existing reactions, is in most cases the same; and further that the nature of this change in the nervous structure is most probably one of increase in the size of the neurone, both with respect to the size of the nerve cell and the cross-section of the axone.

Evidence for these conclusions is drawn from the following five sources: (1) observations by Luciani, Mott, Yates, Langley and Pike and Chappell on recovery of functions following experimental lesions to the nervous system; (2) observations by Warrington and Cajal on the histological changes in nerve cells after experimental lesion; (3) observations by Donaldson and Nagasaki on the increase in the size of nerve fibers and nerve cells towards the close of the period of growth; (4) observations on the structural changes in muscle tissue under conditions similar to those of learning; (5) observations by Rawitz and Langley on the failure of development of the secondary and tertiary afferent neurones, after injury to the primary neurones, in young animals. This evidence is as follows:

a. It has frequently been observed that immediately after injury to the nervous system, certain reactions disappear. In the course of a few weeks many of these may reappear. For example, immediately after spinal hemisection in the cat there is a loss of movement in the homolateral hind limb. After a few weeks the cat will again employ the leg in walking. Two hypotheses have been offered to explain these changes. The immediate loss may be caused by some depressing effect of the injury, such as shock. In the course of time this shock may wear off, permitting the reappearance of the reactions. This hypothesis assumes that the control of the hind limb is a function of the part of the cord below the level of hemi-

section, and as the shock subsides the normal functions of the part reappear. The second hypothesis holds that the loss immediately following the injury is caused by the elimination of some conduction pathways which normally are active in the lost movements. The recovery occurs through an increase in the functional capacity of the pathways which remain intact after the injury. For example, following the spinal hemisection there is an increase in the volume of impulses delivered from the pathways on the uninjured side to the motor cells of the injured side.

The experimental work of Pike and his students has greatly reduced the probability of shock as a possible explanation. They have further shown that recovery of some reactions may occur through an increase in the functional capacity of the conduction pathways on either the afferent or on the efferent side.

It has been shown by Chappell and Pike (2) that splitting the pyramidal decussation of the cat results in some motor limitations immediately after the injury. In a few weeks recovery is so nearly complete that the loss is to be observed with only the most difficult tests. This injury eliminates the crossed pyramidal fibers which descend beyond the medulla. The direct and homolateral pyramidal fibers pass below this level in the cat, but neither constitutes a large tract. If the cells of origin of the pyramidal fibers in the motor area are extirpated after the animal has recovered from the pyramidal decussation injury, the limitations of movement are again precipitated, and to a more marked degree.

As the pyramidal fibers in the cat descend to cross on the ventral aspect of the medulla, they give off collaterals. These pass to the nucleus ruber and other efferent nuclei. That the recovery observed following injury to the pyramidal decussations has occurred through an increase in the functional capacity of the collaterals above the injury seems probable. When the cells of origin are extirpated following the recovery, the loss is quite as severe as it is when the motor cortex is extirpated without any previous injury to the decussation. If the recovery involved any other fibers, this would not be the case.

The question arises as to the probable nature of the changes in the conduction pathways through which this increased functional capacity is mediated. Here again there are two possibilities. Irritability and conductivity are expressions of metabolism. Any change which occurs in this direction must be through the metabolic process. The first possibility is that some change occurs whereby, with the same mass of nervous tissue, an increase in the magnitude of the nervous impulse becomes possible. According to the second a nerve cell is like any other cell in this respect. Its mass, or some part of its mass, changes with the amount of work it is capable of doing.

The probable nature of the structural change underlying compensation for injury to the nervous system is indicated by the observations of Cajal, made in 1911 (1). Using the cells of origin in the cerebellum, he made a unilateral lesion distal to the first collaterals from the axones of these cells. He permitted the animal to recover for a period before making histological examinations of the region. He observed a marked increase in the cross-section of the collaterals on the injured side as compared with those on the uninjured side.

It would appear that increase in the functional capacity of the conduction systems following injury occurs through the second of the above possibilities, that of increased mass.

b. Towards the end of the period of growth, after hyperplasia has ceased in the nervous system, it is observed that the functional capacity of an organism continues to increase. With about the same number of nerve cells, the organism increases the magnitude of its reactions. The observations made by Donaldson and Nagasaki (3) show that hyperplasia continues in the muscular system after it has ceased in the nervous system. As a result an increasing number of muscle fibers must be activated through the same number of nerve fibers. Since it takes work to stimulate muscle fibers, the more muscle fibers that must be stimulated by a nerve fiber, the more work must the nerve fiber do. Its functional capacity increases at least as long as the number of muscle fibers continues to increase. Donaldson and Nagasaki found

that after the close of hyperplasia in the nervous system, hyperplasia in the muscles is attended by hypertrophy in the nerve fibers. They found that not only does the nerve fiber increase in cross-section but that the cell body increases in size. It seems highly probable that in this case, also, the increase in the functional capacity of the conduction pathways is mediated through the observed change in the mass of the neurones.

c. Learning also involves a change in the functional capacity of the conduction pathways. It is observed that before learning, the stimulation of a given group of afferent pathways will not evoke a given reaction. Through the process of learning the reaction does occur when the given afferent group is stimulated. In the learning process there is a change in the functional capacity of the conduction pathways such that they subsequently do more work on a given effector mechanism than they previously did.

Of the two possibilities cited above, which applies to this case? The answer which is commonly given is the first, that metabolic changes involve no change in the mass of the conduction pathways, but may be expressed as a decrease in the synaptic resistance. No relevant evidence can be adduced in support of this belief, but the objections which may be urged against it are fatal. Those who wish to argue for a change in the functional capacity without any change in mass in the conduction system will have to resort to something other than resistance. The available evidence points to a change in mass.

One line of evidence is that taken from muscle tissue under conditions similar to those of learning. If a relatively unused group of muscle fibers is measured and then subjected to the conditions similar to those of learning—that is to say, practice and exercise—in the course of time the circumference of these muscle fibers will be found to have increased. If the subject employed is well beyond the period of adolescence, this increase in circumference must be caused by hypertrophy. The individual fibers of the muscle have increased their cross-section. Stimulation or exercise of the muscle increases its

metabolism. Increased metabolism leads to increase in the cross-section of the muscle fiber. Increase in its cross-section enables the muscle fiber to do more work.

In learning as it is usually conceived we observe only that increase in the excitation or practice leads to the appearance of a response of a given magnitude by stimulation of fewer afferent sources. The internal processes, increased metabolism and a resulting increase in the size of the nerve fiber have not been observed during learning. But since the same general characteristics are found in most tissues, there exists a possibility, and by the method of inductive reasoning even a high probability, that similar changes occur in muscle and nerve under similar conditions.

In the present case it is unnecessary to rely entirely on reasoning of this nature. Support is found in the observations of Warrington (14). After section of the dorsal roots of the spinal nerves, Warrington observed chromatolytic changes in the ventral root cells on the same side. Through this injury, some of the impulses delivered to the ventral horn (motor) cells are eliminated. If the impulses to these cells are still further limited through a hemisection of the spinal cord on the same side and above the level of the section of the dorsal roots, this chromatolytic change is more pronounced. Shäffer has shown that this chromatolytic change attends atrophy. From this, and much other evidence, it is concluded that change in the excitation of nerve cells influences their metabolic activities.

It is further shown by the observations of Rawitz (10) and Langley (7) that excitation affects the growth of nerve cells and fibers. Rawitz found that failure of development of the primary auditory neurones results in failure of development of the secondary and tertiary neurones. This effect may be traced to the temporal lobes. Langley, working with new-born kittens, found that cutting the sciatic nerve resulted in failure of development of the cells and fibers in Clarke's column. These are the cells of origin of the secondary proprioceptor pathway. These observations indicate that for nerve cells and fibers to develop to the size which they com-

monly attain in adult animals, it is necessary that they be excited. Through stimulation the cells and fibers grow larger.

We have the fact that the same conduction pathways do more work, or may react to excitation of fewer peripheral receptors, after learning than before. We have two possible explanations, both of which rest upon a change of some sort in the conduction pathway as an expression of, or result of, a quantitative change in the metabolism in the neurones constituting the pathway. One possibility is that training has induced some change in the nature of the pathway such that it does more work than before, but that there has been no change in its mass. There must, then, be some change in the nature of the reacting mass. The other possibility is that there has been an increase in the reacting mass so that, without any change in nature, the increased mass does more work than before. Comparison with muscle cells and the known changes of mass in nerve cells and fibers during growth make the second possibility the probable one. Any other possibility postulates perpetual motion, which is of doubtful occurrence in living organisms.

II

Summation of Stimuli

The above observations and conclusions are insufficient to explain the manner in which any learned reaction arises. It becomes necessary to consider at some length the manner in which impulses become associated with one another in the nervous system to give rise to reactions. It would appear that association, integration and summation of stimuli all indicate a single central process.

The summation of stimuli in nerve is at once a familiar occurrence in experimental physiology and psychology and a subject in which much mystery still lies concealed under the terminology employed. Any general statement of the problem of summation of stimuli which includes anything more than the summation of successive subliminal stimuli applied at the same point on the nerve seems never to have been presented. Yet many facts have been known, some of them for years, which seem to fit in with the more general

conception of the process of summation of stimuli. It is this more general conception, with some of its specific applications, which will be presented here.

1. The demonstration of the capacity of the nerve fiber to sum up within itself a number of inadequate stimuli is now commonly made by students in the physiology laboratory. The work of Setschenow (**11**), Engelmann (**4**), Lucas (**8**) and many others shows that summation is a property of tissue in general. It is most readily demonstrated in the nerve-muscle preparation of the frog.

If the gastrocnemius muscle and the sciatic nerve of a frog are arranged to record contractions of the muscle, and the stimulating agent is a variable electrical source, a current and potential may be found which will just result in stimulation of the nerve. If now a current of less intensity is employed, it is found that with one application of the stimulus there is no muscular contraction. The stimulus is inadequate to excite the nerve fiber. If, however, this inadequate stimulus is delivered at sufficiently frequent intervals, it will be found, if the stimulus is not too small, that contraction occurs after repeated application. The number of applications required to excite the fiber will vary inversely with the current density and directly with the interval between applications. An inadequate stimulus, though it is incapable of exciting the tissue, does, nevertheless, produce a change in it which persists for a time. A second stimulus is added to the remainder of the change and so on until sufficient change is produced in the nerve to excite it.

The probable nature of this change was explained by Lucas who wrote:

> Hill . . . showed that on Nernst's hypothesis, that the local change induced in a tissue by an exciting current is a concentration of ions, such a type of summation could be calculated; there would be a necessary relation such as I have suggested between the interval necessary for summation and the rate of diffusion of the ions concerned.
>
> Now if this does really represent the mechanism of this summation of subliminal stimuli, it follows that no summation will

be observed unless the exciting action of the two stimuli concerned takes place at the same point on the tissue, for the changes supposed to sum are localized at the seat of excitation. Adrian and I made experiments on this question in the sciatic nerve of the frog and found this inference verified. If the two stimuli fell on points 10 to 15 m.m. apart along the nerve no summation could be observed. There exists then apparently this type of summation which is entirely a local matter, depending on the addition of two incomplete excitatory disturbances at the seat of excitation, and leading to the discharge of a propagated disturbance only when the two local effects have summed to an adequate value (**8**).

2. Passing now to a somewhat different case of summation of stimuli, one may recall the experiments of Stirling (see Sherrington, **12**) and others, upon 'summation at a synapse.' This is not a case of summation of stimuli which are subliminal for the nerve trunk, but the summation of nervous impulses at a synapse, or, as seems more probable, in a nerve cell. A large number of impulses may be necessary to excite a response. Sherrington (**12**) found the scratch reflex occurring after forty excitations of the afferent nerve, while Stirling found that fifty were frequently required. In one case the response occurred after stimulating the afferent root one hundred and twelve times. Here again there is a certain time limit within which the successive impulses must be applied to the nerve if they are to be summed up in the nerve cell. It seems very probable that some change is set up in the nerve cell or dendrites which outlasts the duration of the impulse. It does not seem probable that under ordinary conditions an impulse passing along a nerve is completely blocked at the first synapse it encounters as it is difficult to see why, without the intervention of some entelechy or homunculus, a second impulse would have any better chance of passing than did the first. It seems more probable, and more in line with the facts, that the first impulse gets through the synapse and sets up a change in the nerve cell. A second or third impulse coming along the nerve within a definite time limit would add quantitatively to the change set up by the first and eventually lead to stimulation. Although not usually considered as

belonging in the same category as summation of stimuli in a nerve trunk, it is very possible and even probable that 'summation at a synapse' is analogous in all respects to summation of subliminal stimuli in a nerve trunk. With reference to the secondary cell, single nervous impulses delivered to it are inadequate stimuli, and a number of these single impulses must be summed up in the nerve cell to excite it.

3. It is the common belief that nerve cells are excited either by peripheral stimuli or by impulses reaching the nerve cell from some more peripheral fiber, and by nothing else. While it is known that nerve cells and fibers may be stimulated by a wide variety of agents in the laboratory, the presence of some of these in the organism under biological conditions has received but scant attention.

If a nerve be stimulated by electrical, thermal, mechanical or chemical agents, so far as has been discovered, the nerve impulse set up by any one of these is identical with that set up by another. The nature of the nerve impulse would seem to be independent of the means used to set it up. Furthermore, so far as any one knows at present, the nature of the impulse sent out by a nerve is also independent of the method of stimulation. If, then, some of these agents are present under biological conditions, they must affect nerve cells in the direction of stimulation. Their effect would add to that of impulses coming from the periphery and the result would be analogous in all respects to 'summation at a synapse.' Based on these considerations, an extension of the idea of summation of stimuli may be made to cover certain cases which sometimes have been spoken of as changes in irritability, sometimes as inhibition, sometimes as facilitation and sometimes in more mysterious terms with which the physiological psychologist has little concern. Of the commonly known phenomena of the laboratory, three—(*a*) the effect of the anode and cathode on the excitability of a nerve, (*b*) the effect of deficient circulation on the phrenic nerve, and (*c*) the increased excitability accompanying the early stages of degeneration—seem susceptible to being brought into the category of summation of stimuli in a nerve trunk.

a. When an electrical circuit is closed on a nerve-muscle preparation, stimulation occurs in the region of the nerve which is in contact with the cathode. While the current is flowing at a constant potential, there is commonly no reaction. When the nerve is particularly excitable, one may get an actual tetanus persisting for some time after closure of the circuit (Ritter's tetanus). Ordinarily, however, there is stimulation at the cathode only on closure of the circuit. But a change in the excitability of the nerve persists for a considerable time after closure of the circuit, or as we are inclined to believe, during the entire time of flow of the current. The phenomenon of Ritter's tetanus shows that this change is in the direction of stimulation and that it may even amount to stimulation at times. It seems reasonable to suppose, therefore, that the change persisting during the flow of current is comparable with that set up by a subliminal stimulus, and that as less work is necessary to bring about actual stimulation of the nerve in the region of the cathode, the effect of the stimulating current is merely added to the effect set up about the cathode by the continuous or polarizing current. Since stimulation occurs at the anode on opening the circuit, the increased excitability of the nerve in what is called the post anodic region would become intelligible on the same basis. It is merely another phase of Keith Lucas's period of supernormal excitability. Occasionally, also, one finds that the stimulation which occurs about the anode on opening the circuit persists for some time and gives rise to an actual tetanus of the muscle (Wundt's tetanus). On rare occasions both Wundt's and Ritter's tetanus have been seen in the same preparation.

b. The left phrenic nerve comes in contact with the pericardial sac in its course through the thorax. Students are frequently shown that a nerve of a rheoscopic frog may be stimulated by the action current of the heart or other muscle but they are not told why one is not afflicted with sinistral hiccoughs resulting from the stimulation of the left phrenic by the action current of the heart. Ordinarily the phrenic is not stimulated by the action currents of the heart, nor are

any of the muscles of the forearm stimulated by the action currents of the biceps or triceps, though they pass close to one or another of these. Nor is one fiber of the nerve trunk excited by the passage of nerve impulses in the neighboring fibers; otherwise there would be no law of isolated conduction. But in each of these cases, excitation would have to occur if the only factors involved in the excitation of the nerve of the rheoscopic frog were action currents. Fortunately or unfortunately for organisms, there is another factor, the increase in excitability in the early stages of degeneration. Degeneration does not occur in nerve fibers under biological conditions. It has, however, been observed by Langendorf (**6**), Stewart and Pike (**13**), and Pike and Chappell[1] that conditions may arise under which the left phrenic is stimulated by the action currents of the heart. It is not necessary to sever the phrenic and thereby initiate degeneration in the severed portion, as is done in the frog's sciatic in the demonstration of the rheoscopic frog, nor to invade the thorax in such a way as to affect either the heart or the phrenic nerve directly. Langendorff observed stimulation in the phrenic nerve by action currents of the heart under conditions of hemorrhage and was inclined to believe that under such conditions the heart offered a worse short circuit for its own action currents than it did ordinarily. Stewart and Pike regarded the increased excitability of the phrenic, which often follows ligation of the left subclavian artery below the origin of the vertebral and internal mammary arteries, as a particular instance under the Ritter-Valli law. Subsequently Pike (**9**) expressed the opinion that the accumulation of carbon dioxide in the upper region of the nerve, caused by ligation of the internal mammary, was concerned in this hyper-excitability. The whole matter is so closely bound up with the Ritter-Valli law that some consideration must be given the law.

c. It is well known that in the early stages of degeneration the excitability of the nerve is increased. It is known, also, that this wave of increased excitability begins at the proximal end of the peripheral stump and moves outward as time

[1] Not previously reported.

elapses and the degeneration procedes. This is the Ritter-Valli law, or, as Wilby has phrased it, 'the march of death.' One might say, perhaps, that either the constitution of the nerve fiber is affected in the early stages of degeneration in such a way as to render the nerve more excitable, or that the metabolic changes in degeneration tend in the same direction as those set up during stimulation of the nerve fibers. There is, perhaps, not sufficient evidence at present to enable one to decide which statement is the more correct one, but if it is supposed that the metabolic changes in degeneration tend in the same direction as those occurring in stimulation, the reason for the increased excitability would not be far to seek. If, for example, with the accumulation of carbon dioxide in the upper region of the left phrenic nerve, a degenerative change is initiated, work would be done on the nerve in the direction of stimulation. An alternative view is that carbon dioxide itself does work on the nerve and that, in the range between the minimal and maximal effects of the carbon dioxide, the change in the nerve is in the direction of stimulation. Under these conditions less work would be necessary to complete the change than to start it from the beginning. The action current of the heart might then become sufficient to complete the change. Even at such a time a summation of action currents may be necessary. We have observed in a cat to which a lethal dose of absinthe had been administered, that the diaphragm contracted with every fifth beat of the heart for some minutes previous to death. Similarly in the stimulation of an excised nerve-muscle preparation by action currents, the increased excitability in the region of injury is essential to the production of the response. Provisionally, then, the phenomena of increased excitability comprised or included under the Ritter-Valli law might be regarded as belonging in the same category as the other cases of summation of stimuli.[2]

[2] There exists a possibility that the reactions observed in some psychotic cases commonly regarded as functional, can be understood in terms of the summating effect of degeneration. Thus it may be observed that certain reactions occur, sometimes repeatedly, seemingly without adequate external stimulation. At such a time no cortical degeneration may be found if the case comes to autopsy. If, however, death

4. Passing now to a somewhat different phase of the problem, we may consider the summation of stimuli reaching one and the same nerve cell over different afferent nerves. One may find in the text-books and literature various figures showing the terminations of two or more afferent nerves about the same central cell. It is generally considered that the central cell may be excited by impulses reaching it over either or any of these afferent neurones which form terminations with its dendrites. It has frequently been observed that an animal will react more readily when several peripheral stimuli are applied simultaneously than it will when the reaction occurs in response to one, or even two stimuli. On the basis of the All-or-Nothing law, it is difficult to see how such a result is possible if it is supposed that the cell is stimulated by an impulse delivered to it from any one afferent channel. It is also difficult to see how the result could be obtained unless an impulse reaching a central cell over some afferent fiber was communicated beyond the synaptic membrane to the cell body. But if a nerve impulse reaching the central cell over one afferent fiber is transmitted beyond the synaptic membrane, over the dendrite to the cell body, it may there set up a change in the direction of stimulation and still be inadequate to complete excitation. It is relatively easy to see how another impulse, reaching the same central cell over another afferent fiber and connecting with a wholly different dendrite, may also be transmitted past the synaptic membrane and along a different dendrite to the cell body, and there be added to the change set up by the impulse reaching the cell over the first dendrite. It seems more in accordance with the facts, therefore, to suppose that impulses reaching the same central nerve cell over two or more different afferent routes may set up changes in the central cell which may be added

ensues after dementia becomes profound, autopsy reveals degeneration. Common procedures do not reveal degeneration until the reaction is complete. The 'march of death' may consume considerable time, and during this time, according to the Ritter-Valli law, the metabolic changes will tend in the direction of stimulation. The diseased cells would be excited by stimuli which under conditions of health would be inadequate. Reactions would occur to 'irrelevant' stimuli. We shall consider the general effects of the unchecked activity of the nervous system upon the organism and the fundamental physiological basis of this activity in a future paper.

and thereby result in complete excitation.[8] This enables one to dispense with the postulated but undemonstrated property of the nervous system termed synaptic resistance.

The termination of the end arborizations of several afferent nerves about one and the same central cell is a common condition in the central nervous system, but its real functional significance is less commonly recognized. The idea of summation of impulses over different afferent nerves seems essential to the explanation of the failure of costal respiration after section of the dorsal roots of the thoracic portion of the spinal cord. Costal respiration may proceed unaffected, or only slightly affected after laminectomy and complete exposure of the whole thoracic region of the spinal cord. After incision of the dura, and division of all the dorsal (afferent) roots of the spinal nerves in this region, costal movements of respiration cease altogether or become merely passive so that the ribs sink in with each descent of the diaphragm. If, however, the stomach is filled with water, so that the diaphragm can no longer contract, costal movements of respiration are resumed. There can be, then, no acutal paralysis in the motor neurones or any other neurones of this region of the spinal cord, and no complete blockage of conduction of efferent impulses from the medulla oblongata to these motor cells.

If one supposes that these motor neurones are normally receiving afferent impulses from the muscles of the thorax, in addition to those descending from the medulla, so that summation occurs, one can see how they would respond less readily when the afferent impulses from the intercostal muscles are eliminated. And one can see, also, how, with the accumulation of carbon dioxide in the circulating fluids of the body, following the cessation of the movements of the diaphragm, these motor cells of the spinal cord should finally respond to the impulses coming down from the medulla alone.

The whole appreciation of these considerations will show first, the general insufficiency of evidence for the current idea that wherever one neurone makes synaptic connection with

[8] Herrick, in his discussion of Fig. 40 (14) tacitly assumes some such condition as this in his explanation of the reinforcement of a reflex response.

the next one in a chain, a reaction normally occurs without any associated or adjuvant impulses from any other neurones; second, the quagmire in which one finds oneself at times in attempting to use shock as an explanation of the limitation of movement following injury to the nervous system. One need not postulate shock at all in order to explain why a dog does not use its hind leg immediately after section of the dorsal roots of the lumbar nerves of that side. All that is necessary is to suppose that the motor horn cells normally react under the influence of two or more separate sources—those stimuli coming in over the dorsal roots and making connections through the shorter pathways in the spinal cord, and those coming from the anterior end. After some days have elapsed following the operation, the impulses from the anterior end, in the absence of those from the dorsal roots, are sufficient to excite the cells of the motor horn. These cells also respond when convulsions are set up by absinthe, even immediately after section of the dorsal roots.

Similarly, the maintenance of posture and the Rhomberg sign can be adequately understood only in terms of the summation of impulses from more than one afferent source. A normal individual may stand indefinitely with his feet together and his eyes closed. If, however, as in *tabes dorsalis*, the impulses from the legs are eliminated by dorsal root injury, a subject sways and falls on closing his eyes. At least three groups of impulses enter into the maintenance of posture—visual, vestibular and kinæsthetic. Any one of these may be removed and with the other two intact, posture may be maintained. If a second is eliminated, as when the tabetic closes his eyes, there is not sufficient summation of stimuli and the postural reaction does not appear.

While there are few reactions whose nervous control has ever been completely analyzed, any such analysis progresses but a short way before impulses from two or more afferent sources are found associated with the reaction.

5. It was remarked above that work may be done on the nerve cell by agents other than the nervous impulse. Chemicals in the blood may form a very important group of sum-

mating agents, doing work on the nerve cell in the direction of stimulation and summing up with the effects of nervous impulses to complete excitation. The greater the work that is done by the chemical, the less is that required to be done by the nervous system. If, as is sometimes the case, the chemical does more work and the nervous system does just as much as it did previously, the reaction will be of greater magnitude. The influence of a chemical on the nervous system may be seen in respiration. While there are afferent impulses from at least three sources entering into respiration, there are also chemical impulses involved.

The respiratory rate and the minute volume of respiration, when all the afferent nerves are intact, undergo changes which are related to the concentration of substances dissolved in the blood, particularly carbon dioxide. A change in the afferent impulses reaching the central mechanism over the vagus nerve will likewise bring about a change in the respiratory rate, but without greatly affecting the minute volume. When each of the afferent nerves, so far as they may be reached by operative measures, are cut, the rate of respiration falls greatly. But such measurements as have been made do not show any commensurate change in the minute volume of respiration. If the rate, say twenty a minute when all the nerves are intact, falls to ten a minute without a change in the minute volume after division of both vagi, the obvious inference is that at each inspiration twice as much air is taken into the lungs after section of the vagi as was taken in previously. It would also seem that following bilateral vagotomy the concentration of carbon dioxide must become greater and the concentration of oxygen less, in the interval between successive inspirations.[4] It must follow, therefore, that a greater stimulus is necessary, reckoned in chemical terms alone, to elicit the movement of inspiration after division of both vagi than before.

The question arises as to why such an increased stimulus should be necessary under one set of conditions and not under another. Most of the answers in the literature come perilously

[4] Probably both carbon dioxide and oxygen are of significance in this condition.

close to invoking vitalism and none gives any idea of the process in terms of known reactions in nerve cells. When one considers the fact that stimulation of most of the somatic afferent nerves, such as the brachial or the sciatic, increase the respiratory rate and volume to such an extent that the concentration of carbon dioxide in the blood falls off significantly, it does not seem beyond the range of possibility to suppose that afferent impulses do have a very real relationship to respiratory rhythm.

If one supposes that stimulation of an afferent nerve will in itself affect the respiratory rate by stimulation of the central cells of the medulla, the increase of the minute volume of respiration and the consequent washing out of carbon dioxide in the blood is to be regarded as a reaction of the central cells in accordance with the theorem of LeChatelier. If the cells in the medulla are normally acted upon by two kinds of stimuli, one of which is chemical and the other nervous, it would be expected, first that an increase in the volume of afferent impulses would render unnecessary the usual volume or intensity of the chemical stimuli, second that a decrease in the usual volume of afferent impulses reaching the central mechanism should render necessary an increase in the intensity of chemical stimulation required to maintain respiratory rhythm. The experimental results show that both the above conditions are realized in turn when one brings about the necessary changes in the volume of afferent impulses which reach the central mechanism in the medulla. The conception of summation of stimuli of two different natures, chemical and nervous, in the central cells of the respiratory mechanism affords an adequate explanation of the changes in respiratory rhythm following stimulation of the afferent nerves or elimination by anatomical division or otherwise, of the afferent impulses reaching the central cells.

Summation of stimuli in the nervous system, it seems probable, is a general condition. In fact it is difficult to understand on any other basis how organisms avoid going into convulsions whenever any afferent source is stimulated.

III

An Explanation of Some Varieties of Learning

We have now to examine into the possible application of these considerations to the problem of learning. Briefly stated the two conceptions fundamental to this application are, (1) that, as a result of excitation and practice, nerve fibers and cells grow larger and are thereby able to do more work on the nerve cells to which they are connected, and (2) that in every response afferent impulses from two or more sources are required. If a number of sources are required to produce a response and, through learning, some of the sources can do more work than they did previously, the result will be that fewer sources will be required to initiate a response of a given magnitude.

At first glance this conclusion may seem to be at odds with the All-or-Nothing law. If a nerve or effector fiber responds to its full capacity or not at all, it would seem to follow that by increasing the work done on the cells of origin of the efferent fibers, no increase in the magnitude of the response would be obtained. This assumption, however, omits from consideration the necessary summation of stimuli. When a number of afferent sources are stimulated, the resulting summation on some of the central cells is sufficiently great to produce a given reaction. However, impulses are delivered to many other cells also, but in them the summation is inadequate and they are not stimulated. If the amount of work done on these latter cells is just a little greater, they too will be stimulated and the magnitude of the response will increase. It is well, also, to note that the All-or-Nothing law applies to separate fibers and not to a nerve trunk or effector as a whole.

The possible nature of learning may be more apparent if it is followed through diagrammatically. For this purpose we may use the case of the dog which learns to salivate when a bell is rung. The salivary response is produced, not by stimulation of the tongue alone as is commonly postulated, but by a summation of stimuli which include, or may include, visual, auditory, visceral, tactual, olfactory and a number of

other stimuli, as, for example, chemicals in the body. For the sake of convenience, only three of these are chosen for schematic purposes. To increase the number employed in the diagram above three will merely increase the complexity unnecessarily. Reduction below three does, however, eliminate certain possibilities. It is also to be noted that summation may occur anywhere in the central system where two or more fibers end about the same neurone. The summation is shown only on the efferent nerve cells and this again is in the interest of simplicity.

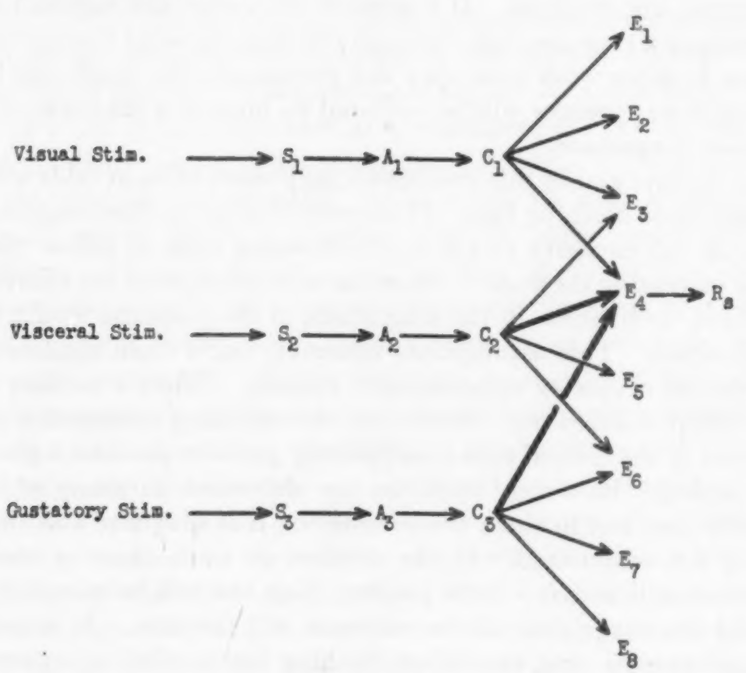


DIAGRAM 1

Assume that a dog is tested under a given intensity of light, with a given degree of starvation and with acid of a given concentration. A salivary response of a given magnitude is elicited. This condition is represented in Diagram 1, in which S indicates an afferent or sensory endorgan, A an afferent pathway, C a central pathway and E an efferent pathway. Summation of stimuli is seen to occur on three

efferent groups but only in the case of E_4 is summation sufficient to lead to a response. The pathways C_2–E_4 and C_3-E_4 are drawn in heavy lines to indicate that these pathways are large and normally do more work in the production of salivation than does C_1–E_4. Consideration shows that these pathways are natively more closely associated with salivation and they also have achieved a maximum of development through practice. They have previously been excited in most of the salivary reactions.

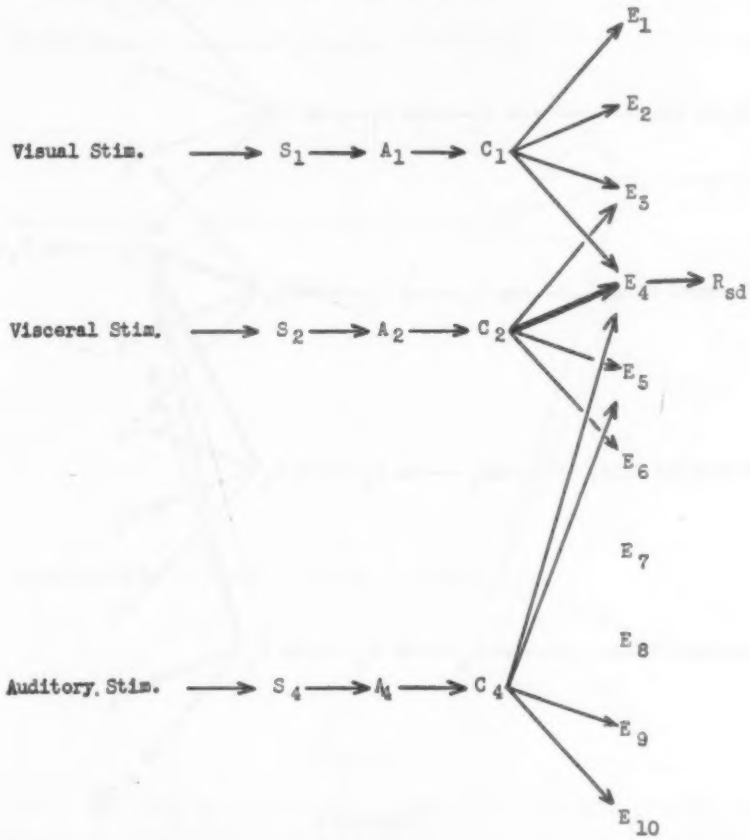


DIAGRAM 2

Under similar visual and visceral conditions, but in the absence of a gustatory stimulus, the addition of an auditory stimulus of a given magnitude does not cause a marked

salivary response. This is represented in Diagram 2. Although there is a nervous pathway between the auditory receptor and the efferent cells in the salivary mechanism, it does little work on the efferent cells and when summed with the impulses from the visual and visceral sources, is still inadequate to stimulate a response of appreciable magnitude in the salivary mechanism. This diminished salivary response is indicated in the diagram by R_{sd}.

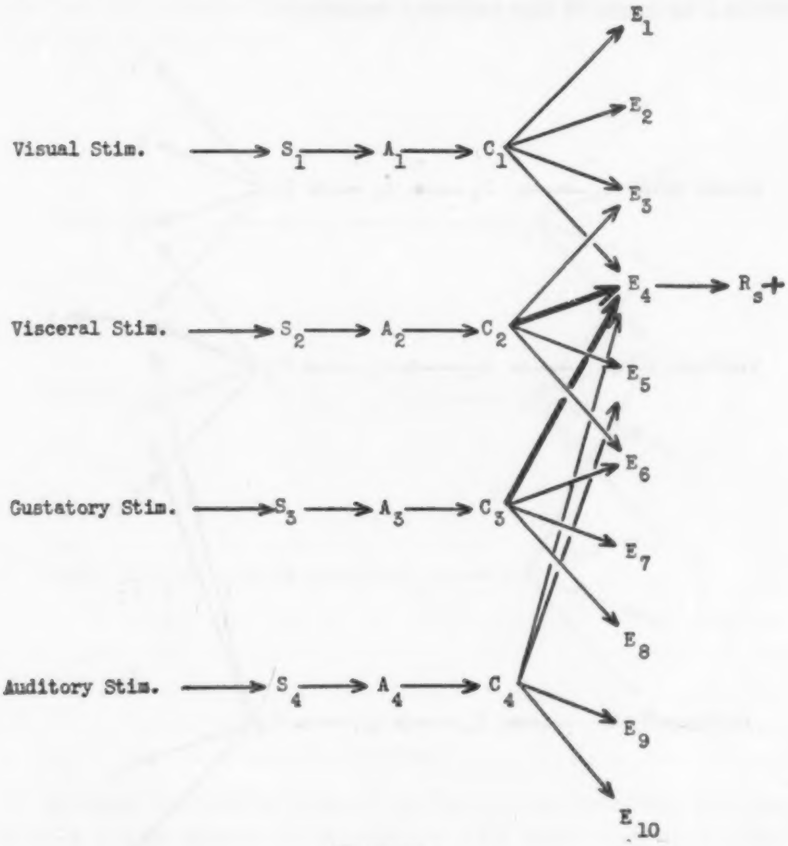


DIAGRAM 3

That some work is done by the auditory mechanism, even without training, is apparent as soon as the learning is undertaken. When the four stimuli are presented simultaneously the salivary response is greater than in the absence

of the auditory stimulus. This is represented in Diagram 3, in which R_s+ indicates an increase in the salivary response above that shown in the first diagram.

If, as this process is repeated, the nerve cells and fibers exercised increase in size, they will become capable of doing more work on those effector cells with which they are con-

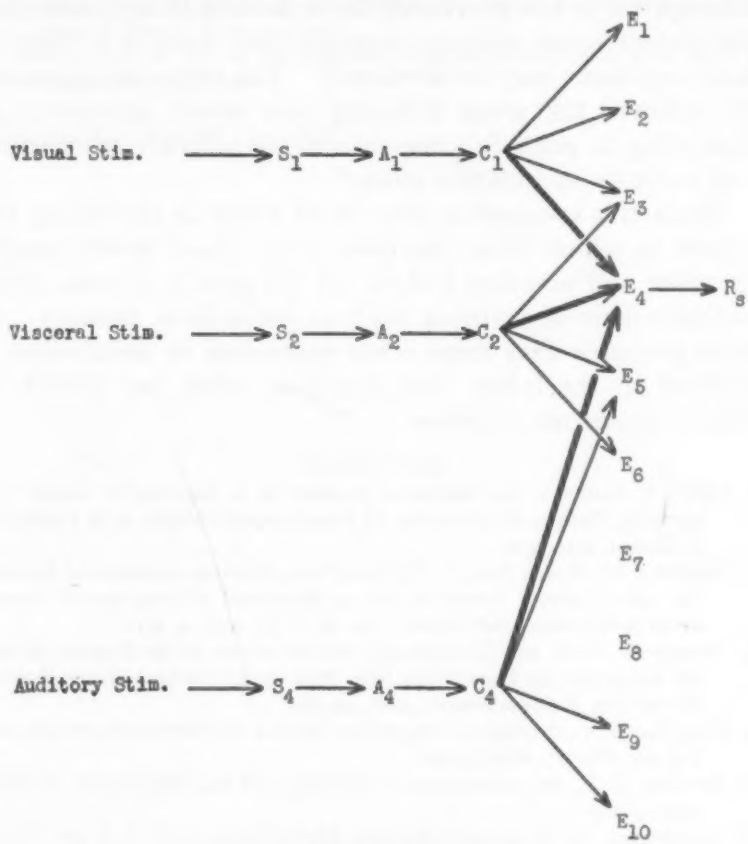


DIAGRAM 4

nected. As the increase continues the condition will be achieved in which the amount of work being done on E_4 by the visual, visceral and auditory pathways will be about equal to that done by the visual, visceral and gustatory pathways before the learning was undertaken. At such a time the gustatory stimulus may be omitted and the response of

salivation will occur. This is shown in Diagram 4. A heavy line indicates that the pathway C_4–E_4 now does more work than it did previously. If the visual connection previously did little work in the production of salivation, the same change must necessarily occur in this pathway. On the other hand it seems probable that no great change will occur in the visceral pathway, for it has previously been developed in connection with almost every salivary response and there is a limit to which any tissue may be developed. The same consideration will indicate the great difficulty one would encounter in attempting to get a full response of the salivary mechanism from auditory stimulation alone.[5]

While this conception may be of value in explaining the manner in which those reactions come about which require repetition and practice, it does not aid greatly in some other manifestations of learning such as immediate memory. It seems probable that some other expression of metabolism is involved in the latter case, but just what, we should, at present, prefer not to guess.

REFERENCES

1. Cajal, S. Ramon y, Los fenómenos precoces de la degeneración neural en el cerebello, *Trabajos del laboratorie del investigaciones biologica de la Universidad de Madrid*, 1911, **9**, 1.
2. Chappell, M. N. and Pike, F. H., The effects of certain experimental lesions of the central nervous system of cats as determined by some specific tests of motor performance, *Bull. Neurol. Inst. of N. Y.*, 1935, **4**, 451.
3. Donaldson, H. H. and Nagasaki, G., On the increase in the diameter of nerve cell-bodies and the fibers arising from them during the later phases of growth (albino rat), *J. Comp. Neurol.*, 1918, **29**, 529.
4. Engelmann, W., Beiträge zur allgemeinen Muskel und Nerven-physiologie, *Arch. f. d. ges. Physiol.*, 1870, **3**, 248.
5. Herrick, C. J., An introduction to neurology, 3rd ed., Philadelphia: Saunders, 1922, p. 107.
6. Langendorf, O., Elektrophysiologische Mittheilungen, *Arch. f. d. ges. Physiol.*, 1903, **43**, 277.
7. Langley, J. N., On the degeneration resulting from removal of the cerebral cortex and corpora striata in the dog, *J. Physiol.*, 1890, **11**, 606.
8. Lucas, K., Conduction of the nervous impulse, Monographs of Physiology, London, 1917, p. 23.
9. Pike, F. H., A note on the excitation of the phrenic nerve by the action current of the heart, *Amer. J. Physiol.*, 1907, **19**, 339.

[5] This conception has previously been applied, in part, to the explanation of the 'inhibition' of a conditioned response, Psychol. Rev., 1931, **38**, 317.

10. Rawitz, B., Gehörorgan und Gehirn eines wiessen Hundes mit blauen Augen, *Schwalbe Morphol. Arb.*, 1896, **6**, 545.
11. Setschenow, J., Physiologische Studien über die Hemmungsmechanismen für die Reflextatigkeit des Ruckenmarks im Gehirne des Frosches., Berlin, 1863 (bound pamphlets).
12. Sherrington, C. S., Integrative action of the nervous system, New Haven: Yale Univ. Press, 1920, p. 20.
13. Stewart, G. N. and Pike, F. H., Resuscitation of the respiratory and other bulbar nervous mechanisms with special reference to the question of their automaticity, *Amer. J. Physiol.*, 1907, **19**, 339.
14. Warrington, W. B., On the structural alterations observed in nerve cells, *J. Physiol.*, 1898, **23**, 112.

[MS. received August 12, 1935]

INTELLIGENCE AND THE HUMAN BRAIN

BY JAMES A. HAMILTON

Psychological Laboratory, University of California

In an attempt to formulate a plan for the systematic investigation of the brain of two strains of rats a considerable number of studies of the human brain were consulted. Certain unique features of the literature were immediately apparent and at the same time certain intriguing questions arose.

It is beyond the purview of this paper to investigate all of the 335 publications which have been examined.[1] It must suffice to point out only the most important and influential papers and to indicate trends, hypotheses and sources of error rather than specific details. A peculiar characteristic of the literature is the tenacity of several hypotheses, based on deduction rather than experimental grounds, which resisted the most destructive laboratory evidence. Investigators spent years only to disprove what had been disproved a quarter of a century before. Even more frequently an experimenter failed to avoid an experimental hazard, the existence of which had been suggested long before his time. A large portion of the disappointments which were suffered and the mistakes which were made can be attributed to ignorance or disregard of the findings of others.

The nature of the historical material to be presented does not lend itself to logical organization. Those who studied the brain itself searched for the characteristics of genius first in gross size, then in complexity of fissuration, then in configuration, and more recently in the vascular system of the brain. Despite the intimate theoretical association between studies of brain size and of skull size, the cranial investigations were carried on by an almost entirely independent group. For

[1] A complete bibliography is on file in the Library of the University of California. The writer is indebted to Miss Fern Scneder for valuable assistance in the accumulation and examination of these publications.

this reason it would seem advisable to follow the history of brain studies through to the present before examining the data from cranial measurements.

Probably the first suggestion that the depth and complexity of the fissures of the brain are related to intelligence is credited by Galen (**13**) to Erasistratus. The latter is said to have asserted that the greater complexity of the human brain, as compared with that of lower animals, was attributable to the higher mental powers of the humans. Galen took exception to this opinion, and suggested that if Erasistratus were correct, one would expect the ass, because of its stupidity, to have an extremely smooth and simple brain.

The brains of three great men, Cromwell, Byron and Cuvier, were extremely important in establishing the concept, almost universal in the first half of the nineteenth century, that intelligence was closely related to brain size. In 1860 Rudolph Wagner (**34**, no. 7) was struck with the discrepancy between the accepted figures for these three brains and the brains of five of his Göttingen colleagues. The former ranged from 1829 to 2238 grams, while the latter group ranged from 1226 to 1520 grams. Wagner had a Dr. Southard investigate the sources of the data, in the case of the brains of Cromwell, Byron and Cuvier, and the results of this investigation were published shortly thereafter (**34**, no. 12). The figure for Cromwell's brain was traced to a remark in a Latin commentary published in 1702, forty-two years after Cromwell's death, and was subordinate to a statement regarding the condition of his bowels and liver at autopsy. The weight of Byron's brain was originally given as an even six pounds, but since the autopsy was performed in Italy, it was impossible to determine whether this weight was in Neopolitan, Venetian or English pounds. The unit of measurement selected might change the weight by 400 grams. It is interesting to note that in 1931, or 71 years after Wagner's exposé, the unrevised weights of the Cromwell and Byron brains are cited by Buytendijk (**8**).

The various reports of the weight of Cuvier's brain differed by as much as one pound and eight ounces, and there was

evidence of abnormal condition of the brain at autopsy. Furthermore, one of Cuvier's physicians contributed the interesting fact that Cuvier had had tendencies toward hydrocephaly in his youth, and that almost all of his children had died of this ailment.

Frederick Tiedemann's comparison of the Negro, European and Orang-outang brains, published in 1836 in the 'Philosophical Transactions of London' was frequently quoted by nineteenth century writers. Tiedemann is credited by Bain (1) in 1864 with establishing "an indisputable connexion between the size of brain and mental energy displayed by the individual man," largely by finding three microcephalic brains which weighed 560, 730 and 638 grams respectively.

The year 1860 was marked by Wagner's (**35**) important introduction to the morphology of the human brain. In addition to a detailed study of the brains of four scientists and one laborer, Wagner presented a table giving a comprehensive summary of data on 964 brains as reported by Virchow, Rousseau, Bergmann, Tiedemann, Parchappe, Huschke and Sims. Almost all of these brains were from insane persons, bandits and paupers. The heaviest brain of all was that of an hydrocephalic idiot. Somewhat disconcerted by the fact that his scholars were far from the top of this group, Wagner strongly criticized current assumptions regarding the functional importance of gross brain size, and formulated a stirring plea for immediate investigation of the problem, and specifically for contributions of brain specimens from individuals of superior intellect.

Wagner's paper met with immediate and widespread response. Almost the entire volume of the 1861 'Memoirs of the Societé d'Anthropologie de Paris' was devoted to a discussion of the human brain. This volume contains Broca's (**6**) series of brain measurements, which were of sufficient importance to induce Topinard to revise and reclassify them 20 years later. It is interesting to note that Broca was unable to discover any connection between brain size and intelligence; indeed, he ventured to opine that such persons as Cuvier may have achieved distinction in spite of, rather than because of

the magnitude of their brains. In 1885 Topinard (**33**) was able to report that 10,182 brains had been weighed by 35 investigators. A few of these men are of special importance because of the accuracy of their work, the number of brains examined and the fact that their measurements became the standards with which later brains were compared. The most important were Wagner (**35**), Broca (**6**), Bischoff (**5**) and Marchand (**22**). It should be noted that each of these men took pains to mention that, in his own series, there was no evidence of an association between size of brain and intellectual endowment.

A second type of response to Wagner's plea was the formation of societies for the purpose of accumulating large collections of the 'brains of the élite' (**30**). Members of the societies signed mutual wills regarding the disposal of their respective brains. The first of these organizations was the Societé d'Autopsie de Paris, which was founded in 1881 as a subsidiary group within the Societé d'Anthropologie. It was followed in 1889 by the American Anthropometric Association, and shortly thereafter by the Cornell Brain Association, under the leadership of Professor Burt G. Wilder.

As the 'brains of the élite' began to accumulate, professional journals became filled with papers devoted to the description of isolated brains of men of genius. These papers invariably included the weights of the brains, although in many cases the brains had been weighed after, rather than before fixation. In most cases the brains were described minutely, and elaborate tables of measurements were usually added. Beautiful etchings and engravings of the brains were not uncommon.

For the moment let us consider only such data from these papers as are related to the weight of the brains. Spitzka (**30**) has assembled data from more than one hundred élite brains as reported in over one hundred and fifty publications. The distribution of weights in one hundred cases is compared with a distribution of 1334 weights from the series of Bischoff, Marchand and Topinard. The mean of the élite series is 1471 grams, while that of the 'normal' series is 1364 grams. There

is thus a difference of 107 grams (.60 sigma of the 'normals') in favor of the élite group. While such a difference is not particularly significant, even it must be examined with a degree of skepticism.

It should be noted that material of Broca (from which Topinard's tables were constructed), Bischoff and Marchand was taken largely from cases in public institutions, particularly hospitals and institutions for the feebleminded and insane. Among such cases it is more than likely that there were not a few persons characterized by microcephaly, and perhaps others with degenerative conditions of the brain. The sample cannot be justly termed 'normal' in any modern sense of the word.

The technical skill employed and the uniformity of conditions adhered to by such scientists as Broca, Bischoff and Marchand cannot be questioned. Each of these men describes in detail his own procedure. A comparison of methods used reveals no important dissimilarity. Unfortunately, there is no such assurance in the case of many of the investigators of individual brains whose reports constitute the raw data for Spitzka's table. Few of the investigators of brains of the élite gave data regarding the details of preparation; probably many were unaware of the hazards of such work. Some idea of the factors affecting brain weight may be obtained from the study of the brain of H. L. F. von Helmholtz by Hansemann (**16**). The brain as removed from the cranium weighed 1700 grams. This would place Helmholtz in the ninety-third percentile of Spitzka's list of the élite. The blood coagulum was then removed, which reduced the weight to 1540 grams (seventy-third percentile). So much blood still remained on the brain that Hansemann estimated that the 'real' weight of the brain was from 1420 to 1440 grams (forty-fourth percentile). This single test of the reliability of measurement might well excite our suspicion of the studies which contain no report of ambiguities of measurement.

There is a selective factor which is likely to influence the distribution of data such as that presented by Spitzka. The general tenor of most papers indicates that the expectation of

most investigators was to find the brain of a man of genius characterized by large size and complex convolutions. It is probable that many autopsies revealed brains which were not outstanding in any way. It is likely that many such findings remained unpublished; such an outcome is not uncommon in the general run of research. Only an unusual man such as Wilder (**37**) is likely to publish an account of an 'intelligent white man' who had a brain weighing but 680 grams, or less than half the usual endowment.

It would seem that the early expectations regarding the intimate relation of gross brain size to intelligence were not borne out in the experimental laboratory. Closer examination of the evidence casts grave doubt on the validity of any assumption of relationship based solely on studies of the brains of men of genius. The difficulties confronting comparison of brain weights, however, are minor when compared with the difficulties involved in the comparison of descriptions of brains. The earlier investigators expected to find (as did Erasistratus 2000 years before them) deep fissures and many convolutions in the brains of the élite. Such expectations were upset by Professor Wilder (**36**), who said of the brain of the noted mathematician and philosopher Chauncey Wright, "The simplicity of the fissures, and the width and flatness of the gyres are paralleled in the Cornell collection only in the much smaller brain of an unknown mulatto. Some approach to this condition occurs in Ruloff, a murderer, and perhaps in a German shown by Wagner."

Since there were no accepted methods of description and measurement, adjectives and superlatives were often combined with anatomical designations. With the increase of knowledge of brain function, investigators of élite brains began to discover, in individual brains, particularly well-developed areas which they attempted to correlate with special abilities of their subjects. An area was discovered in the brain of the astronomer Glydén which seemed likely to be associated with mathematical ability. As late as 1929 Papez (**23**), in describing the brain of Professor B. G. Wilder, said, "My own observations are limited to the Cornell collection, but they

lead me to place emphasis more especially on the development of the inferior frontal region surrounding the inferior frontal sulcus, its anterior transverse limb, and the sulcus radiatus in front of it. The superior portions of the frontal region do not seem to exhibit this correlation to the same extent. One can single out such brains as Ernst Haeckel, described by Maurer, and of Helen M. Gardner in support of this view."

Spitzka examined the brain of Joseph Leidy and became convinced that men of genius were distinguished by large size of the corpus callosum (**30**). Bean (**2**) had already completed a study of 150 brains which indicated that such a conclusion was rather precipitate. Fortunately, there were a few dissenters to the general trend of the literature. Steida (**31**) examined the brain of a man who had spoken fifty languages with fluency, and found nothing remarkable in size or configuration. Mall (**21**) criticized the use of a single brain as the basis for conclusions and attempted to use a system of ratings on a series of brains. He found it almost impossible to distinguish the sex of his subjects, much less their traits of personality or intellect.

In contrast with almost all of the recent investigators of anatomical correlates of intelligence is Henry Donaldson, who has the temerity to test his hypotheses under experimentally controlled conditions. His interest dates from 1890 (**10**) and has continued unabated until the present time. His most recent attempt at a solution is of particular interest, since it indicates some of the hazards which confront investigators of individual differences in the brain. Donaldson's (**11**) theory was that since, in the embryo, the anterior and posterior poles of the brain develop before the region between them, one might look for an excess of frontal and occipital regions in the brains of geniuses. His earlier study reports detailed measurements of the brains of G. Stanley Hall, Sir William Osler and Edward S. Morse. The actual measurements were made by Dr. Myrtelle M. Canavan, according to a technique worked out by Donaldson. The frontal and occipital areas of all three scholars were more extensive than those found by Dr. Canavan (**9**) in the case of three other brains measured by her.

These findings suggested the necessity for a large control group, so Donaldson (12) supervised the measurement of brains of 30 white males and 30 negro males. The actual measurement, in this case, was done by Dr. Addison. When the results from the Normal controls were compared with those previously found, it was apparent that the three scholars had a marked deficiency in parietal and an excess of temporal area. In order to assure himself that the differences found were real, Donaldson had Dr. Addison remeasure the brains previously measured by Canavan. The proportional distribution of cortical areas in the case of the scholars was found to be identical with that in the case of the white and negro controls. The differences found had apparently been due to differences in technique, despite the fact that in each case the same system was followed and the work was supervised by the same person.

It is difficult to refute, on the basis of factual data, the assertions of any one investigator, or of all the investigators, who claim to have discovered relationships between traits of intellect and personality and the configuration of individual brains. Each study constitutes a single case which cannot be duplicated. On the other hand, there is abundant historical evidence that the customary procedure was to anticipate such relationships and to 'discover' characteristics of the brain which conformed to the anticipation. If such relationships actually existed, it would seem likely that at least some of the findings might have been duplicated. Yet Spitzka, despite his obvious predisposition, was unable to find any conformity in the description of the brains of men of genius. Moreover, the failure of genuine experimenters, such as Henry Donaldson, whose approach was always sympathetic albeit critical, is significant.

Within the last decade the Russian anatomist Hindzé has suggested a new approach to the problem of anatomical correlates of intelligence. In one publication (17) he shows sections from the vascular tree of the pia mater of the mathematician P. Nekrassoff and compares them with the vascular tree of a bandit and an unknown hospital patient. Differ-

ences in complexity are very apparent. Donaldson (12) has suggested that this approach may be very significant. It is important to note, however, that Hindzé has reported specifically on less than ten cases. No quantitative method has been developed for measuring the degree of complexity of the vascular system of the pia mater. It would seem best to reserve judgment of Hindzé's theory until additional investigations, using adequate measurements and large numbers of cases, are available.

The accumulation and study of human brains, other than those from institutional patients, is a slow process depending both upon the demise of the subject and the bequest of his brain. Moreover, the complicating factor of age is not subject to experimental control. The desirability of using cranial measurements as an index of the configuration and size of the brain was realized early in the nineteenth century. The classic monograph of Gall (14) was instrumental in suggesting that both size and shape of the cranium were intimately related to intelligence and personality. The reader is familiar with the rise and fall of the peculiar type of phrenology which attempted to relate the shape of the skull to specific traits of character. Gall's importance, however, in stimulating research both in the field of cranial measurement and cortical localization, should not be underestimated.

The more recent literature regarding the association between cranial size and intelligence has been ably summarized by Paterson (24). The present paper will include only a few historical matters which are less accessible, together with a consideration of several recent investigations, hypotheses and evaluations.

The 1861 volume of the 'Memoirs of the Societé d'Anthropologie' included numerous suggestions regarding cranial measurements, one of which was elaborated at length by Broca (7). At the meeting of the English Anthropological Institute on April 24, 1888, a paper was read by Dr. A. Venn describing preliminary results from anthropometric measurements taken on Cambridge students (38). There is no record of the publication of this paper. On the following day Francis

Galton (**15**), who was then president of the Institute, commented on Dr. Venn's study and presented a rearrangement of the Cambridge data. Measurements of height, width and length of skull were multiplied together to give an approximate index of the cranial capacities of the men. Galton concluded, "Accepting these curves as a true statement of the case—and they are beyond a doubt an approximately true statement—we find that a 'high honor' man possesses at the age of nineteen a distinctly larger brain than a 'poll' man in the proportion of 241 to 230.5, or one that is almost 5 per cent larger."

Binet's contribution to cranial measurement began in 1898 with an historical review (**3**). His experimental investigations covered a period from 1900 to 1910, and were dominated by an interest in establishing limiting boundaries (*frontières anthropomètriques*) for the diagnosis of feeblemindedness in children. Following much the same statistical procedure as was later used by him in intelligence testing, he established norms for stature, five cranial measurements, for children from the age of 6 to 18 (**4**). He suggested that a retardation of six years or more in cranial development was significant evidence of abnormality, while a retardation of from three to six years, if confirmed by inferior scholarship or unsatisfactory scores on psychological examinations, would be valuable diagnostic evidence.

Contemporary with the French investigations were studies by German, English and American investigators. The development of rating methods and intelligence testing, and the availability of large numbers of cases made studies of this nature particularly adaptable to statistical analysis. Pearson (**26**), using four-gold tables, correlated breadth of head with scholastic ability in the case of 524 Cambridge students, and found the r to be $\pm$ .0843 $\pm$ .024. He averaged correlations for 5000 pupils of various school grades and concluded that the mean r of both length and breadth of head with scholastic ability was of the order of $+$ .065.

Pearl (**25**) applied the method of mean square contingency to the data of Loewenfeld and Eyerich (**20**) on 935 Bavarian

soldiers rated into four classes by their superiors, and found an r of $+.14 - .04$ between head circumference and intelligence. His conclusion is an unusually astute analysis of the problem: "If further statistics (of which there is great need) should show that generally there is a just sensible positive correlation between these characters, the correct interpretation of the fact would, it seems to me, probably be physiologic rather than psychologic. That is to say, the association between vigor in growth (leading to a well-developed body) and vigor in mental processes would most probably be the result of the action of good conditions of nurture. Other things being equal groups of men with well-nourished bodies are on the average likely to be more able intellectually than groups in which bad conditions of nutrition prevail. Such an interpretation of the facts seems at present to have a much better justification than any which in effect implies that a big brain connotes *per se* an abie mind. Rightly interpreted the facts regarding the correlation between size of head and intelligence, seem to me to furnish, so far as they go, direct statistical evidence in favor of the adage: *mens sana in corpore sano.*"

Examination of the literature summarized by Paterson (**24**) and others reveals that almost all the correlations between head measurements and intelligence are of the order of from $+.05$ to $+.10$. One or two exceptional studies report correlations as high as $+.20$. In striking contrast to the great mass of data which has accumulated are two coefficients reported respectively by Porteus and Berry (**27**) and Hull (**18**). The former study reports a correlation of $+.30$ between intelligence test scores and head measurements, but fails to give the slightest indication of the number of subjects or of the nature of the intelligence tests used.

Hull applied a multiple regression technique to data taken from the A. B. thesis of Elsie B. Sherman and found that a combination of three head measurements and two facial angles correlated with scholarship to the extent of $+.50$. He points out that such a correlation is of the order of magnitude usually found between scholarship and intelligence test scores, and suggests that some sort of cranial index might be valuable in the prognosis of scholastic ability.

The writer has consulted the original manuscript of Sherman's thesis (**28**). Her subjects were 78 freshman students of engineering at the University of Wisconsin, and were of German, Chinese, Norwegian, Russian, Bohemian and American stock. Ages ranged from 16 to 34. Such an heterogeneity of sample suggests that the multiple r which was obtained may be unique for the group which was under consideration, and that measurements of a second population, similarly weighted, might well result in a negligible or even zero order correlation.

Summary

Studies of the human brain have given us no assurance that an association exists between special or general intellectual ability and the size or configuration of the brain, in a normal or superior sample of the population. Measurements of the cranium correlate with intelligence test scores, school grades or ratings almost uniformly between + .05 and + .10. Since the size of the cranium is to some extent related to the size of the brain (**19**), it is likely that an adequately controlled study would reveal some small degree of correlation between brain size and intelligence. The writer would conjecture that such a correlation would not excede + .15. Even if a correlation of this magnitude were obtained, there would remain the possibility that it represented nothing more than a reflection of the small positive correlation which exists between almost all desirable traits.

Many of the investigations described in this paper represent an eminently sane, reasonable application of logical principles. If the brain, the organ of mind, ascends the phylogenetic scale hand in hand with intellectual ability, it would seem most likely that individual differences in intelligence within a single species might be determined largely by size of the brain. If there is a roughly specific localization of function within the several lobes of the brain, it is apparent that special development of a particular area should be paralleled with unusual talent associated with that area. Such principles, however, do not fit experimental findings.

Individual differences in intelligence may be determined by one or two anatomical factors which may some day be isolated, or they may be determined by such a multiplicity of factors as to preclude any simple solution of the problem. If we continue to search for large determining factors, it may be wise to abandon all considerations of reason or logic. Possibly Hindzé's work is a step in this direction. The only other suggestion, as yet untested in the laboratory, is that of Simkins (29), who adjures us to look to the supragranular cortex if we would solve the problems of genius and mediocrity.

REFERENCES

1. Bain, A., The senses and the intellect, London: Longman Green, 2nd ed., 1864, pp. 12–13.
2. Bean, R. B., Some racial peculiarities of the Negro brain, *Amer. J. Anat.*, 1906, 5, 353–360.
3. Binet, A., Historique des recherches sur les rapports de l'intelligence avec la grandeur et la forme de la tête, *Année Psychol.*, 1898, 5, 245–298.
4. ——, Les signes physiques de l'intelligence chez les enfants, *Année Psychol.*, 1910, 16, 1–30.
5. Bischoff, T. L. W. von, Das Hirngewicht des Menschen, Bonn: P. Neusser, 1880, pp. 171 and tables.
6. Broca, P., Sur le volume et la forme du cerveau, *Bull. Soc. Anthrop.*, 1861, 2, 139–207.
7. ——, Sur le craniograph et sur la détermination de plusieurs angles nouveaux nommés angles auriculaires, *Bull. Soc. Anthrop.*, 1861, 2, 673–680.
8. Buytendijk, F. J. J., Le cerveau et l'intelligence, *J. de Psychol.*, 1931, 28, 345–371.
9. Canavan, M. M., Elmer Ernest Southard and his parents, Cambridge, Mass.: University Press, 1925.
10. Donaldson, H. H., Anatomical observations on the brain and several sense-organs of the blind deaf-mute, Laura Dewey Bridgman, *Amer. J. Psychol.*, 1890, 4, 1–98.
11. Donaldson, H. H., A Study of the brains of three scholars: Granville Stanley Hall, Sir William Osler and Edward Sylvester Morse, *J. Comp. Neurol.*, 1928, 46, 1–95.
12. Donaldson, H. H., The brain problem, *Amer. J. Psychiat.*, 1932, 12, 199–214.
13. Galenii, Claudii, De usu partium corporis humani, Lib. VIII, Cap. 13. In Kühn's Medicorum graecorum opera, Vol. III, pp. 672–674. Leipzig, 1822.
14. Gall, F. J., Anatomie et physiologie du système nerveux en général et du cerveau en particulier, Paris, 1810–1819, 4 vol.
15. Galton, F., Head growth in students at the University of Cambridge, *Nature*, 1888, 38, 14–15.
16. Hansemann, D., Über das Gehirn von Hermann v. Helmholtz, *Zeit. f. Psychol. u. Physiol.*, 1899, 20, 13–26.
17. Hindzé, B. K., Wie sollen wir die Hirnarterien verarbeiten? Zur methodik der makroskopischen Erforschung der Hirnarterien, *Ziet. f. d. ges. Neurol. u. Psychiat.*, 1931, 132, 443–455.

18. Hull, C. L., Aptitude testing, Yonkers: World Book, 1928, 131–138.
19. Lee, A., Study of the correlation of the human skull, *Science*, 1900, **12**, 946–949.
20. Loewenfeld, L., and Eyerich, G., Über die Beziehung des Kopfumfanges zur Körperlänge und zur geistigen Entwicklung, Wiesbaden: J. F. Bergmann, 1906, pp. 1–55.
21. Mall, F. P., On several anatomical characters of the human brain, *Amer. J. Anat.*, 1909, **9**, 1–32.
22. Marchand, F., Ueber das Hirngewicht des Menschen, *Abh. d. math-phys. Cl. d. Königl. Sächsischen Ges. d. Wiss.*, 1902, **27**, 391–482.
23. Papez, J. W., The brain of Burt Green Wilder, 1841–1925, *J. Comp. Neurol.*, 1929, **47**, 285–341.
24. Paterson, D. G., Physique and intellect, New York: Century, 1930, pp. 304.
25. Pearl, R., On the correlation between intelligence and the size of the head, *J. Comp. Neurol. and Psychol.*, 1906, **16**, 189–199.
26. Pearson, K., On the correlation of intellectual ability with the size and shape of the head, *Proc. Royal Soc.*, 1901, **69**, 333–342.
27. Porteus, S. D., and Babcock, M. E., Temperament and race, Boston: Gorham Press, 1926, p. 149.
28. Sherman, E. B., An experimental investigation concerning correlations between certain head measurements and University grades. Thesis submitted for degree of B.A., University of Wisconsin, 1923. Ms.
29. Simkins, C. S., The physical basis of intelligence, *Sci. Mo.*, 1930, **31**, 519–523.
30. Spitzka, E. A., A study of the brains of six eminent scientists and scholars belonging to the American Anthropometric Society, together with a description of the skull of Prof. E. D. Cope, *Trans. Amer. Phil. Soc.*, 1907, **21**, 175–308.
31. Steida, L., Das Gehirn eines Sprachkundigen, *Zeit. f. Morph. u. Anthropol.*, 1907, **11**, 83–138.
32. Topinard, P., Le poids du cerveau d'après les registres de Paul Broca, *Rev. d'Anthrop.*, 1882, **5**, 1–30.
33. ——, Eléments d'anthropologie générale, Paris: Adrien Delahaye, 1885, pp. 628.
34. Wagner, R., Kritische und experimentelle Untersuchungen über die Funktionen des Gehirns, *Nachr. d. Königl. Ges. d. Wiss. zu Gött.*, 1860, No. 7, Feb. 29, and No. 12, Apr. 16.
35. ——, Vorstudien zu einer wissenschaftlichen Morphologie und Physiologie des menschlichen Gehirns als Seelenorgane, mit besonderer Rüchsicht auf die Hirnbildung intelligenter Männer, *Abh. d. Königl. Ges. d. Wiss. zu Gött.*, 1860, **1**, 59–152.
36. Wilder, B. G., The cerebral fissures of two philosophers, Chauncey Wright and James Edward Oliver, *J. Comp. Neurol.*, 1895, **5**, 124–125.
37. ——, A brain of about one half the average weight from an intelligent white man, *Proc. Amer. Phil. Soc.*, 1910, **49**, 188–190.
38. News item: *Nature*, 1888, **38**, 23.

[MS. received September 3, 1935]

THE WEBER-RATIO FOR INTENSITIVE DISCRIMINATION *

BY ALFRED H. HOLWAY AND CARROLL C. PRATT

Harvard University

It is now over one hundred years since E. H. Weber (41, 137) first stated in effect that the ratio $\Delta R/R$ (a relative measure of a just discriminable difference) is independent of the magnitude of R (the measure employed to estimate a magnitude of the 'stimulus') for a given sensory mechanism. Figure 1 expresses Weber's belief graphically. The results of

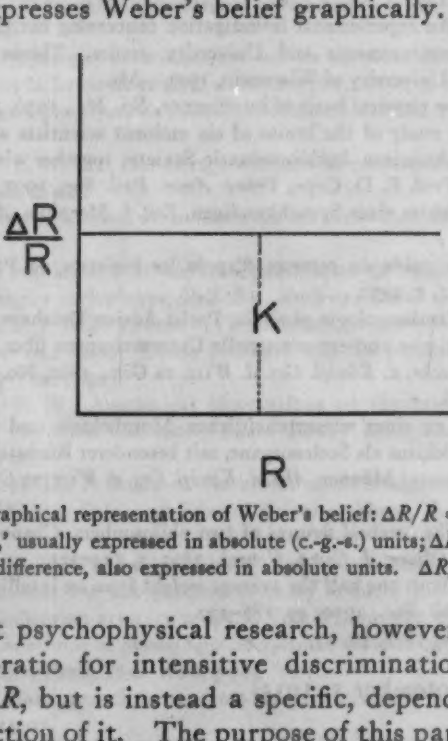


FIG. 1. Graphical representation of Weber's belief: $\Delta R/R = k$. R = a measure of the 'stimulus,' usually expressed in absolute (c.-g.-s.) units; ΔR = a measure of the just noticeable difference, also expressed in absolute units. $\Delta R/R$ is a pure number.

more recent psychophysical research, however, indicate that the Weber-ratio for intensitive discrimination is not independent of R, but is instead a specific, dependent and reproducible function of it. The purpose of this paper is to exhibit graphically, and then to discuss, the special nature of this dependency of $\Delta R/R$ upon R for particular sensory systems.

* With special reference to Fechner's 'Fundamental Value' of the stimulus.

Vision—In 1888, inspired by Fechner's 'Fundamentalformel' and its theoretical implications, Arthur König and Eugen Brodhun (**24**, **25**) began a series of exceptionally complete studies of the Weber-fraction and its relation to the intensity of visual stimuli. Taking special precautions to avoid qualitative modulation, using two subjects and the method of limits (just noticeable differences), these workers determined the values of $\Delta R/R$ for white light throughout a range of more than nine logarithmic units of R (**24**, 643).

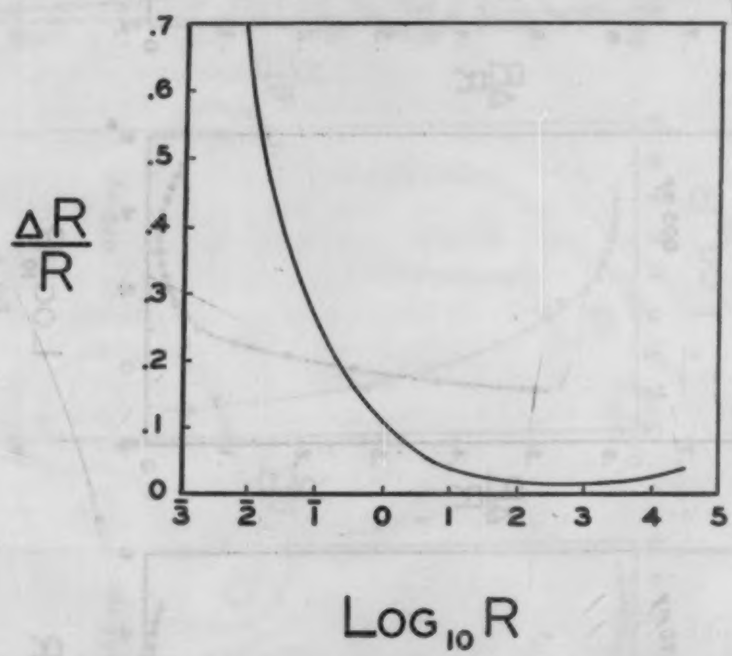


Fig. 2. *White light.* Weber-ratio as a function of log R, expressed in photons. The ordinates are averages for the eyes of König and Brodhun, as plotted by Boring.

Hecht (**18**, **19**, **20**) has compared and indicated the close degree of correspondence between these results and those obtained by Aubert (**1**) and Blanchard (**3**). Hecht's R-values are expressed in millilamberts. Troland (**38**, 78) has averaged these results and converted the abscissæ into log-photons. Boring (**8**, 200) averaged the results obtained by König and Brodhun, and also expressed the abscissæ in terms of log-

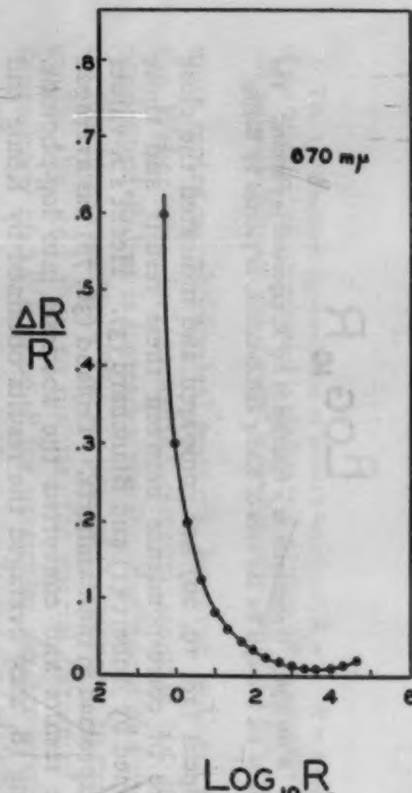


FIG. 3.

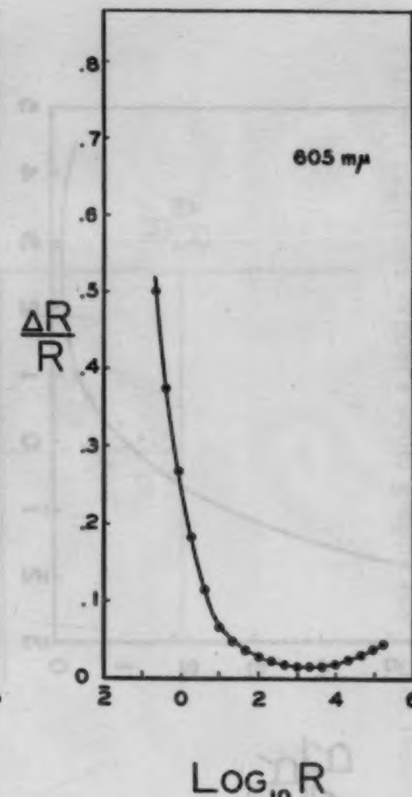


FIG. 4.

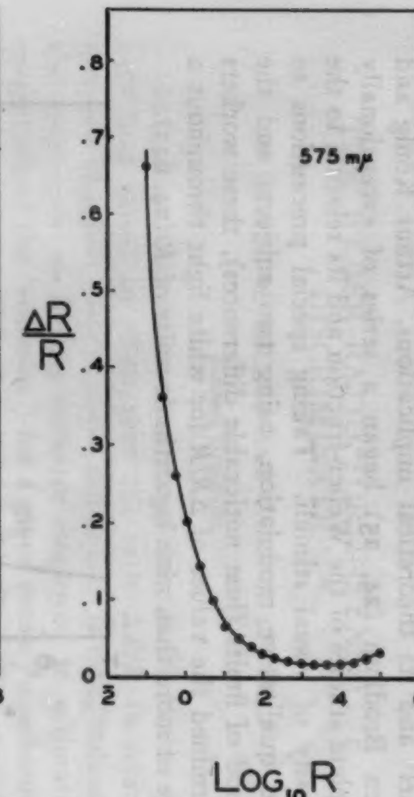


FIG. 5.

FIGS. 3, 4, and 5. *Chroma.* Weber-ratios, for three different wave-lengths, as functions of log *R* in meter-candles. Data for König's eye.

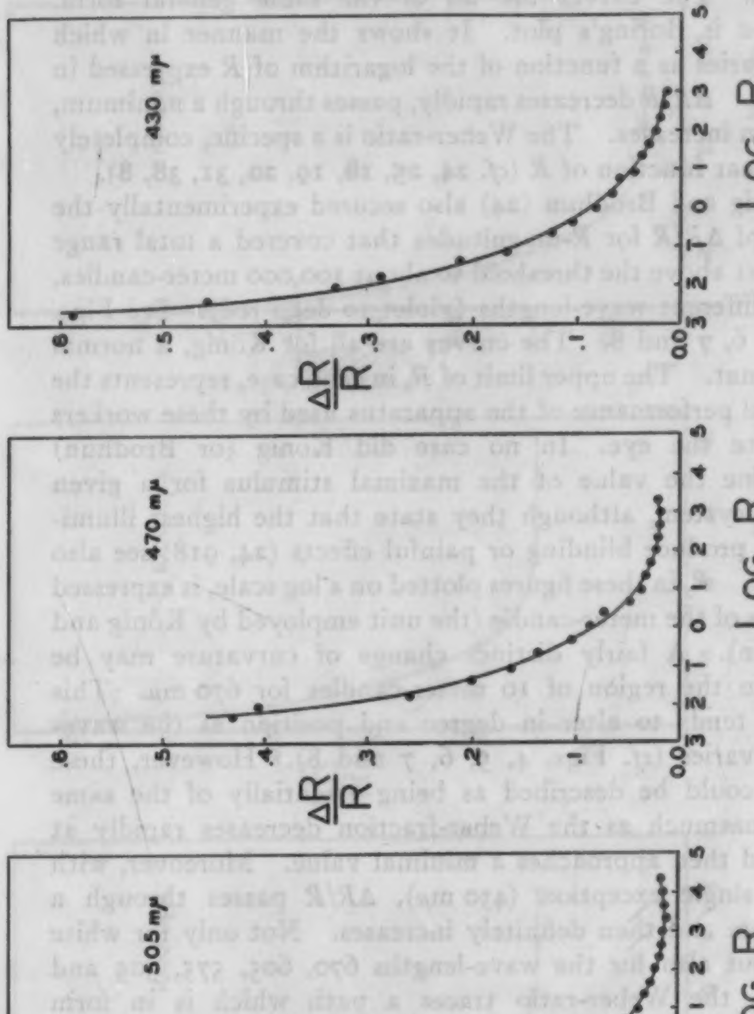


FIG. 6. FIG. 7. FIG. 8.

FIGS. 6, 7, and 8. *Chroma.* Weber-ratios, for three different wave-lengths, as functions of log R in meter-candles.
Data for König's eye.

photons. The curves are all of the same general form. Figure 2 is Boring's plot. It shows the manner in which $\Delta R/R$ varies as a function of the logarithm of R expressed in photons. $\Delta R/R$ decreases rapidly, passes through a minimum, and then increases. The Weber-ratio is a specific, completely curvilinear function of R (*cf.* **24**, **25**, **18**, **19**, **20**, **31**, **38**, **8**).

König and Brodhun (**24**) also secured experimentally the values of $\Delta R/R$ for R-magnitudes that covered a total range from just above the threshold to about 100,000 meter-candles, at six different wave-lengths (violet to deep red). See Figs. 3, 4, 5, 6, 7 and 8. The curves are all for König, a normal trichromat. The upper limit of R, in each case, represents the maximal performance of the apparatus used by these workers to excite the eye. In no case did König (or Brodhun) determine the value of the maximal stimulus for a given sensory system, although they state that the highest illuminations produce blinding or painful effects (**24**, 918; see also **19**, 763). R, in these figures plotted on a log scale, is expressed in terms of the meter-candle (the unit employed by König and Brodhun). A fairly distinct change of curvature may be noted in the region of 10 meter-candles for 670 mμ. This change tends to alter in degree and position as the wave-length varies (*cf.* Figs. 4, 5, 6, 7 and 8). However, these curves could be described as being essentially of the same type, inasmuch as the Weber-fraction decreases rapidly at first and then approaches a minimal value. Moreover, with but a single exception (430 mμ), $\Delta R/R$ passes through a minimum and then definitely increases. Not only for white light, but also for the wave-lengths 670, 605, 575, 505 and 470 mμ, the Weber-ratio traces a path which is in form essentially identical with that depicted in Fig. 2.

Tone—Figures 9, 10 and 11 show $\Delta R/R$ plots for tonal intensity. The data for Fig. 9 are due to Knudsen (**23**). Each point on the curve is an average of 24 judgments (method of limits) for one subject. Two tones (1,000 c.-p.-s.) of equal duration were alternated (rate = approximately 50 per min.) at a sound source which was usually a telephone receiver actuated by energy from a vacuum tube oscillator. Report of

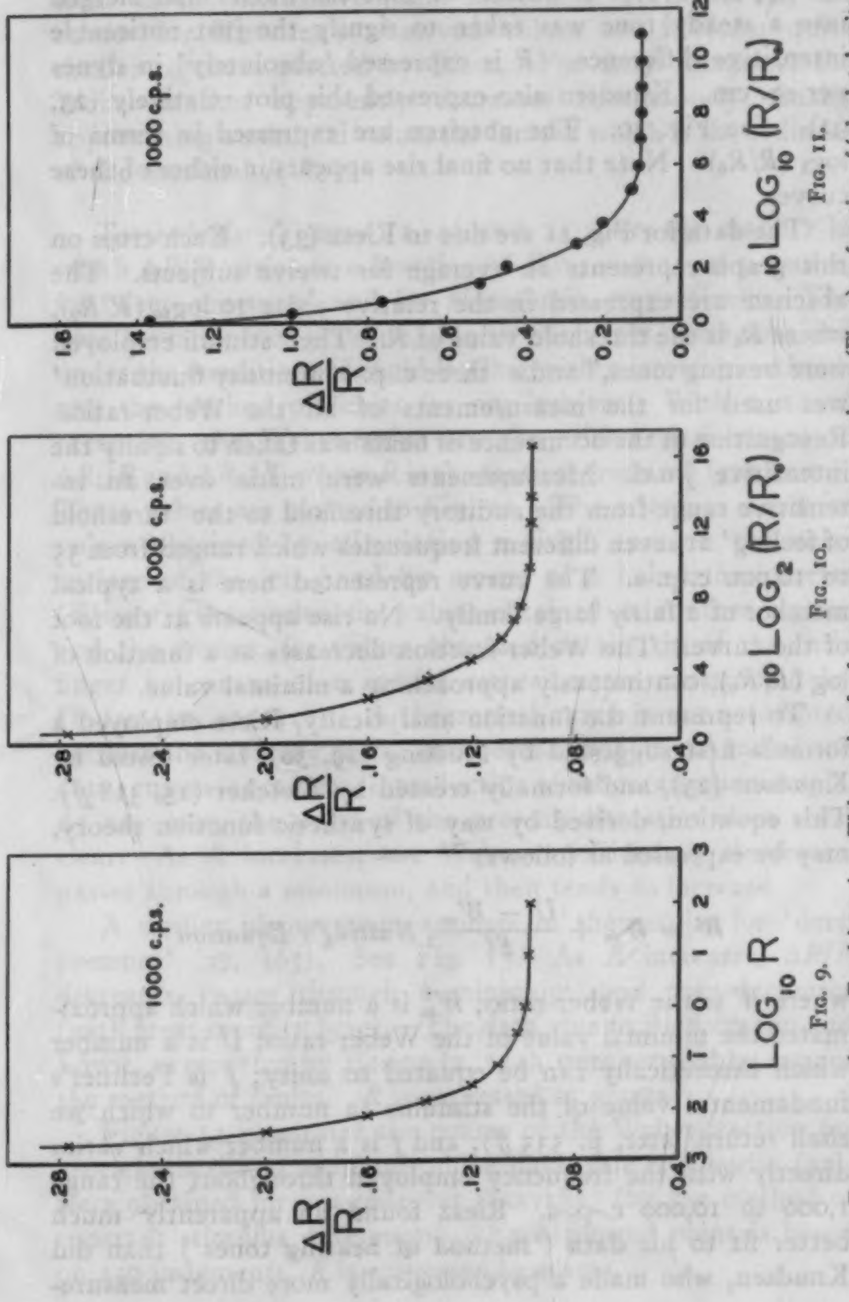


FIG. 9. FIG. 10. FIG. 11.

FIGS. 9, 10, and 11. *Tonal intensity.* Weber-ratios, for frequency = 1,000 c.-p.-s. shown (Fig. 9) as function of 'absolute' units and (Figs. 10 and 11) as function of relative units. Data estimated with aid of reading lens from graphs presented by Knudsen and by Riesz. See text.

the appearance of a 'flutter' or that the flutter had merged into a steady tone was taken to signify the just noticeable intensitive difference. R is expressed 'absolutely' in dynes per sq. cm. Knudsen also expressed this plot relatively (**23**, 92). See Fig. 10. The abscissæ are expressed in terms of $\log_2 (R/R_0)$. Note that no final rise appears in either of these curves.

The data for Fig. 11 are due to Riesz (**33**). Each cross on this graph represents an average for twelve subjects. The abscissæ are expressed in the relative units $10 \log_{10} (R/R_0)$, where R_0 is the threshold value of R. The "stimuli employed were beating tones," and a 'three c.-p.-s. intensity fluctuation' was used for the measurements of all the Weber-ratios. Recognition of the occurrence of beats was taken to signify the intensitive j.n.d. Measurements were made over an intensitive range from the auditory threshold to the 'threshold of feeling' at seven different frequencies which ranged from 35 to 10,000 c.-p.-s. The curve represented here is a typical member of a fairly large family. No rise appears at the foot of the curve. The Weber-fraction decreases as a function of $\log (R/R_0)$, continuously approaching a minimal value.

To represent the function analytically, Riesz employed a formula first suggested by Nutting (**29**, **30**), later tested by Knudsen (**23**), and formally treated by Fletcher (**15**, 318 *ff*). This equation, derived by way of synthetic function theory, may be expressed as follows:

$$W = W_m + \frac{U - W_m}{F^f}, \textit{ Nutting's Equation}$$

where W is the Weber-ratio; W_m is a number which approximates the minimal value of the Weber-ratio; U is a number which theoretically can be equated to unity; F is Fechner's fundamental value of the stimulus (a number to which we shall return later, p. 335 *ff*); and f is a number which varies directly with the frequency employed throughout the range 1,000 to 10,000 c.-p.-s. Riesz found an apparently much better fit to his data ('method of beating tones') than did Knudsen, who made a psychologically more direct measure-

ment of the intensitive j.n.d. for tones. Nevertheless, the curves for Knudsen's observed values are in form quite similar to those plotted by Riesz. For tonal intensity, the Weber-fraction apparently decreases continuously, finally approaching a minimal value, the limit of which is the ordinate of the maximal R/R_0.

Somesthesis—Figures 12, 13 and 14 show the manner in which $\Delta R/R$ varies as a function of R for cutaneous pressure, for 'deep pressure,' and for kinesthesis, respectively. The data for cutaneous pressure, due to Gatti (**17**), were obtained under the conditions of rapid loading with microscopic stimuli and the method of limits for one subject. With a minor exception, the Weber-ratios are the arithmetical means of $\Delta R_l/R$ and $\Delta R_u/R$, where R is the tension produced (gm./mm.). These values are plotted in Fig. 12. The triangles stand for values obtained by stimulating a single spot on the volar surface of the left hand by means of a hair-esthesiometer (Kiesow); the circles, for values obtained at another region; and the crosses, for values obtained on the tip of the little finger by means of a special esthesiometer constructed by Dodge (**17**, 423–426). A theoretical curve is not attempted because the meaning of the unexpected drop at the end of one of the curves is not clear (introduction of subcutaneous tissue?). At any rate, the point of the present discussion does seem clear. As R increases, the Weber-ratio initially decreases, passes through a minimum, and then tends to increase.

A similar phenomenon appears in the results for 'deep pressure' (**27**, 363). See Fig. 13. As R increases, $\Delta R/R$ decreases, passes through a minimum, and then increases (with great rapidity here). The data, due to Biedermann and Löwit, as reported by Hering (**2**, 344), were secured by way of the method of limits. R is expressed in grams.

Figure 14 illustrates the course of the Weber-fraction for kinesthesis (lifted weights). The data, due to Oberlin (**32**), were obtained for one observer (Stavsky) (by the method of constant stimulus differences). Each plotted point is based on 250 judgments, R is expressed in grams.

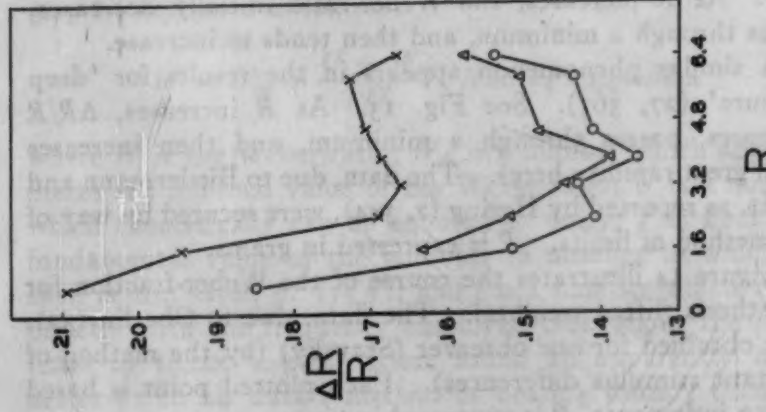


FIG. 12.

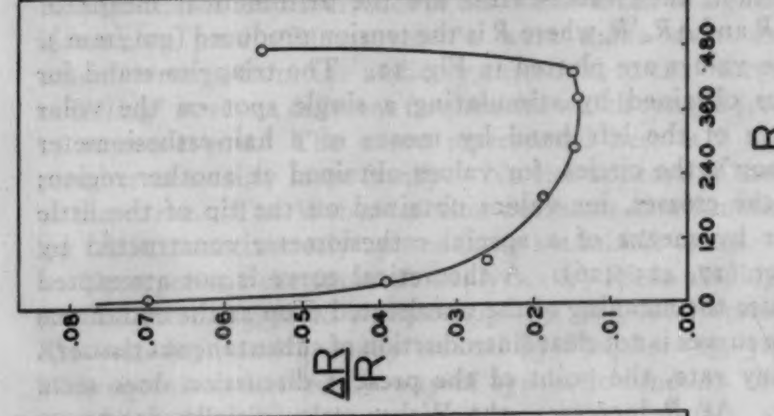


FIG. 13.

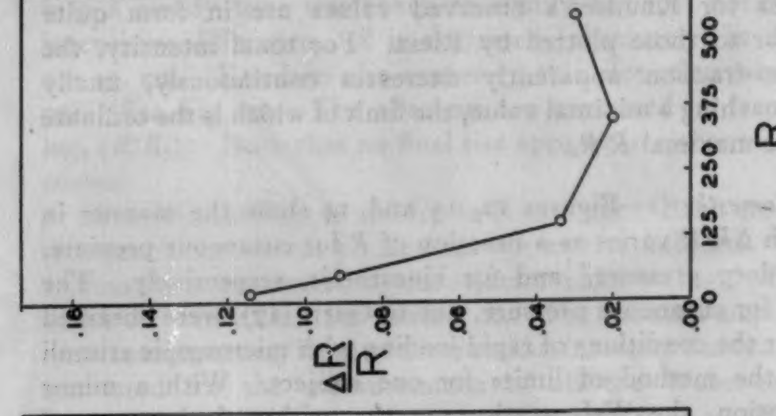
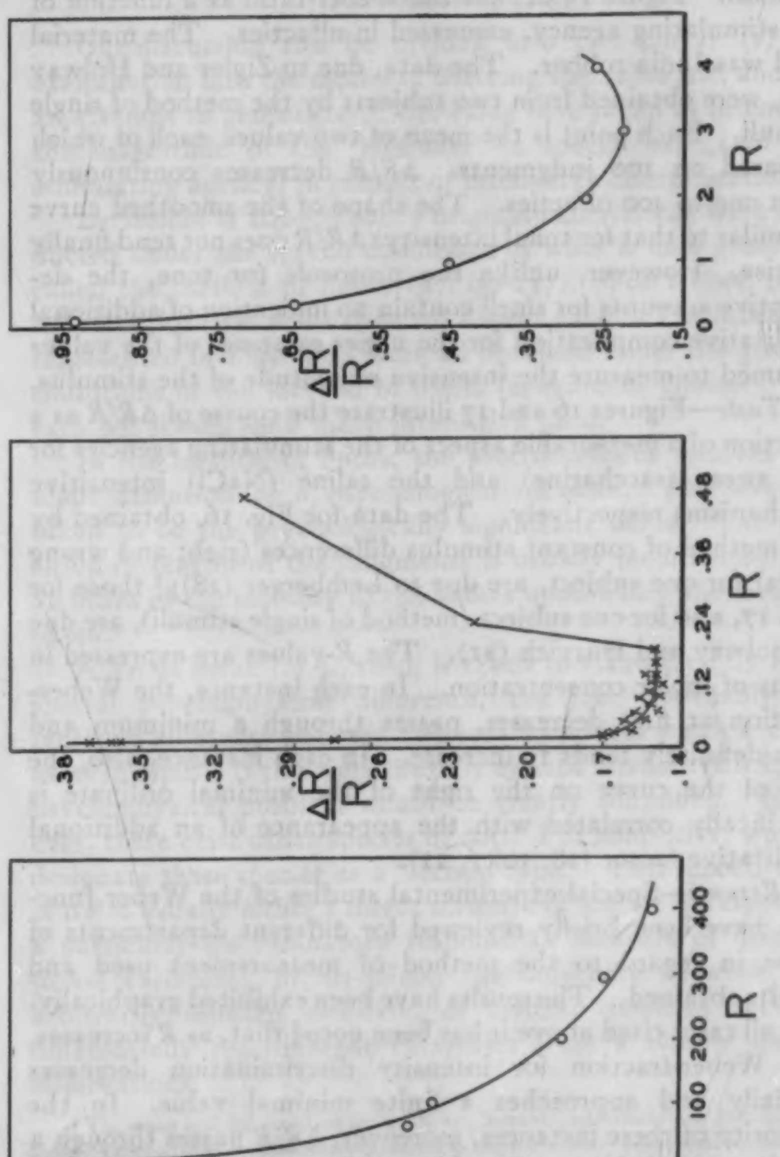


FIG. 14.

FIGS. 12, 13, and 14. *Somethesis.* Weber-ratios expressed as functions of *R* for cutaneous pressure, for 'deep pressure,' and for kinesthesis. See text.



FIGS. 15, 16, and 17. *Smell and taste.* Weber-ratios shown as functions of R. For smell, R is expressed in terms of olfacties; for taste, in terms of molar concentration. See text.

Smell—Figure 15 depicts the Weber-ratio as a function of the stimulating agency, expressed in olfacties. The material used was India rubber. The data, due to Zigler and Holway (**43**), were obtained from two subjects by the method of single stimuli. Each point is the mean of two values, each of which is based on 100 judgments. $\Delta R/R$ decreases continuously from one to 400 olfacties. The shape of the smoothed curve is similar to that for tonal intensity: $\Delta R/R$ does not tend finally to rise. However, unlike the protocols for tone, the descriptive accounts for smell contain no indication of additional qualitative complication for the upper extremes of the values assumed to measure the intensive magnitude of the stimulus.

Taste—Figures 16 and 17 illustrate the course of $\Delta R/R$ as a function of a measurable aspect of the stimulating agencies for the sweet (saccharine) and the saline (NaCl) intensitive mechanisms respectively. The data for Fig. 16, obtained by the method of constant stimulus differences (right and wrong cases) for one subject, are due to Lemberger (**28**);[1] those for Fig. 17, also for one subject (method of single stimuli), are due to Holway and Hurvich (**21**). The R-values are expressed in terms of molar concentration. In each instance, the Weber-fraction at first decreases, passes through a minimum and then definitely tends to increase. In each instance, also, the rise of the curve on the right of the minimal ordinate is specifically correlated with the appearance of an additional qualitative factor (**28**, 308*f*; **21**).

Resumé—Special experimental studies of the Weber function have been briefly reviewed for different departments of sense in regard to the method of measurement used and results obtained. The results have been exhibited graphically. For all cases cited above it has been noted that, as R increases, the Weber-fraction for intensity discrimination decreases initially and approaches a finite minimal value. In the majority of these instances, moreover, $\Delta R/R$ passes through a minimum and then tends to rise. The striking uniformity of these experimentally determined facts presents an invitation to scientific discussion.

[1] Two of Lemberger's values have been omitted for practical considerations. See **28**, 304.

Discussion

Our discussion can be divided into two parts: (1) an examination into the modes of deriving the value ΔR; and (2) an attempt to generalize R, the value here taken as intending the magnitude of the 'stimulus' (a special aspect of the stimulating agency) in respect of intensitive discrimination.

Derivation of ΔR—Weber's original experiments were conducted under the special conditions of what is now generally called the method of limits (**37**, 99–137). For Weber, then, ΔR stood for (was equivalent to) a *mean* j.n.d. The functions represented in Figs. 2–11 were all obtained under the general conditions of the method of limits (continuous change), and each ΔR stands for a mean intensitive j.n.d.

In the method of limits, the precision-index receives but scant attention as a psychological measure. The *mean* is taken to be the psychologically significant value. The precision or scatter of the judgments is usually recorded only as an index of the stability of the values upon which the mean is based.[2]

A ΔR as mean j.n.d., then, is taken to stand directly for a special psychophysical difference, the just noticeably intensitive difference between two sensory effects which are the same in kind. If this were the only type of intensitive ΔR, the psychophysical problem would be greatly simplified. However, there exist other species of ΔR. For simplicity, we can designate these species as a 'second type.' This second type of ΔR is usually either a direct measure of scatter in respect of a particular discriminatory response (a measure of dispositional variation), or, in effect, the difference between two arbitrarily selected estimates of central tendency secured differentially for different categories from a sample set of observations.

[2] The mean and the standard deviation are, however, experimentally suitable and logically adequate as *supplementary* measures in respect of many behavioral phenomena—since, strictly speaking, they are wholly independent of (1) each other and (2) the form of the distribution of the individual measures from which they are derived. For a clear account of an exact relationship which obtains between the mean and standard deviation for visual discrimination, see Crozier (**11**); for proprioceptive adjustment, see (**9**, 39–49).

Obviously, such distinctly different types of ΔR need not vary directly with one another. It is, indeed, a far cry "from the principle of the j.n.d. to the principle of the psychometric function" (7, 1).

On the other hand, if we knew the exact nature of the function relating the two 'types' of ΔR, we should then be able easily to pass from the one to the other; but the function is not known. Although several experimenters have labored with the problem, none has ever been able to penetrate to its roots. No complete account of the relation between ΔR as mean j.n.d. and ΔR as index of dispositional variation has ever been made out even for a single sensory system.[3]

Nevertheless, if a given disposition is forced experimentally to vary in accordance with a special criterion, say in accordance with the phi-function of gamma, then the root-mean-square-deviation of the differential distribution of these variations may alter directly (as does ΔR as mean j.n.d.) with the magnitude of R.[4] At any rate, this is usually the psychological basis of comparison for methodological studies in psychophysics (*cf.*, *e.g.*, **36**, **39**, **14**, **22**). It is also the working hypothesis sometimes adopted by those who employ the method of constant stimuli (single stimuli, etc.) or the method of average error in order ultimately to estimate the relative magnitude of particular discriminatory effects (**28**, **12**, **32**, **43**, **21**).

The ΔR's basic to the functions represented in Figs. 14–17 are of this second type. They are not values which may be taken to stand directly for mean j.n.d.'s. Rather, they are of the order of *variation-indices*. The two ought never to be confused. Fineness of discrimination is by no means the same thing as discriminatory precision. With few exceptions experimental knowledge warrants only the assumption that both types of ΔR vary directly with R, and hence with each other. Thus we are at the present time prevented from attempting more than a gross generalization in regard to the

[3] This deficiency is especially unfortunate because there are cases (*e.g.*, taste and smell) where the method of limits is not conveniently applicable.

[4] It seems probable that, under proper conditions, the interval of uncertainty also varies directly with R; *cf.* (32).

form of the Weber-function from data which depend on these two types of ΔR.

Generalization of R—As heretofore employed, the term R has not been generally defined: photons, meter-candles, millilamberts; dynes per centimeter squared, watts; inches squared, grams, and grams per millimeter; molar concentration. Here we come face to face with the reason why attempts to 'generalize' the relationship between ΔR and R, or between $\Delta R/R$ and R, have always failed to meet with scientific approval: such a feat is physically impossible. Because there is at present no way of knowing the physical relationship between, say, 'a molar concentration' and 'an inch,' these absolute units may not be plotted on the same abscissæ. And this, of course, is what is meant when it is said that R has never been generalized (4; 5, 60 *f*).

However, the recognition of this difficulty does not mean that we are forced to wait and hope for discoveries in the field of physics and chemistry. In spite of the fact that R itself cannot be generalized, there exists a way by means of which we may satisfy the essential requirement of those who have sought this generalization. This amounts to the improvising, on a relative basis, of a common frame of reference for the magnitude R (estimated in absolute units) and its differential effect upon a given sensory system (estimated in the same units): *i.e.*, the introduction of a common or neutral basis for relating the various and physically diverse measures presumed to estimate the 'external' physical and chemical correlates of the attributes (intensity, extensity, etc.) relating to particular sensory events. Although we cannot generalize in terms of absolute (c.-g.-s.) units, we may nevertheless do so in terms of relative units. What is needed is not an absolute but a *relative* system of reference.

Of the several ways in which this end might be attained, we need consider but two:

I. If we knew exactly the total R-range from the stimulus-limen (RL, or R_0) to the maximal stimulus (R_m: *cf.*, the threshold of 'feeling' in audition) for a given sensory system, a relative scale could readily be conceived in terms of per-

centages (*i.e.*, of *ratios*). R_0 would then correspond to zero per cent; R_m, to 100 per cent; and any number lying within this range would be an index of the relative magnitude of R. In addition, this abscissae scale might be utilized as a kind of null or neutral dimension, the units of which would be independent of the absolute difference between R_0 and R_m, and therefore independent of the c.-g.-s. units actually employed in the measurement of R. In terms of this relative dimension, the various physical dimensions of R (mass, length, time) might be referred to the same abscissa.[5]

Attractive as this particular system of reference may appear, however, the fact remains that suitable and sufficient experimental data do not exist to render it practicable at the present time. What is needed, here and now, is a dimension that is experimentally less exacting but which admits of the same type of comparison.

II. Seventy-five years ago, G. T. Fechner set forth a concept which suggests both this very dimension and the unit of it (13, 13 *ff*): *the fundamental value of the stimulus.* This value is, it may be recalled, the ratio of R to its threshold value (R_0) for any particular sensory system. In every case, the unit is the neutral or natural one provided by the stimulus limen in relation to R. This 'fundamental' dimension and its relativistic unit are fully in harmony with the conditions which promise to satisfy our present need.[6]

Figure 18 depicts the behavior of the Weber-ratio as a function of the logarithm of what Fechner called the 'fundamental value of the stimulus' (F), for five different sense-departments.[7] These functions are of neutral, or 'zero,'

[5] The writers are indebted to Professor W. J. Crozier for these suggestions.

For a method of measuring the magnitude of sensory effects strictly in terms of conventional physical dimensions, see Holway and Hurvich (21).

[6] In passing, we should note that Fechner was not alone in recognizing the comparative usefulness of such a numerical system of reference. Zwaardemaker (45, 18 *ff*) utilized the notion when he defined the *olfactie*. Gamble (16), working in Titchener's laboratory, employed the same concept in 1898. Nutting had a similar view (30). Fletcher (15) must indeed have entertained this principle when he adopted the technological *decibel*. And of course there were others: *e.g.*, see (27, 374; 6, 280; 23; 33; 43).

[7] The abscissæ are here expressed in terms of the *logarithm* of F, not in order to adjust the abscissa to the initial intensitive threshold, but merely for visual convenience.

dimensions throughout; that is to say, they are independent of the particular units actually employed in the measurement of the magnitudes involved.

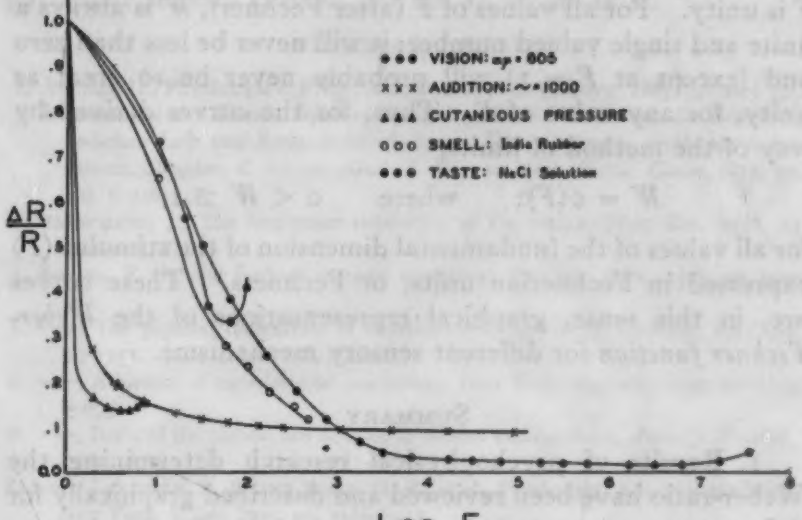


FIG. 18. *Weber-Fechner functions* selected from five different departments of sense. Weber-ratios shown as functions of the logarithm of Fechner's fundamental value of the stimulus.

The data for three of these functions (vision, audition and cutaneous pressure) were secured by the method of limits. The stimulus-thresholds have been experimentally determined for vision and tone. For cutaneous pressure, R_0 is here assumed to be 0.8 gm./mm. for the values represented by circles in Fig. 12 (*cf.* 17, 406, Table 1, Column *b*). These three curves illustrate, relatively, the differential performance of three particular sensory systems which pertain to three distinctly different sense-departments. Functions for the gustatory and olfactory mechanisms are also shown in the figure.[8]

The general assumption has been made that ΔR is equivalent to R at the stimulus-threshold, and $\Delta R/R$ has accordingly

[8] The data for these two systems have been adjusted to the point (1, 1). It should be recalled that the ΔR's basic to these two functions are of the order of variation-indices (*vide supra*).

been set equal to unity as the ordinate corresponding to $F = 1$. For ΔR as mean j.n.d., the Weber-ratio $\left(\frac{R_2 - R_1}{R_2} = \frac{\Delta R}{R}\right)$ which may be called W (after Weber), is equal to unity when F is unity. For all values of F (after Fechner), W is always a finite and single valued number: it will never be less than zero and (except at $F = 1$) will probably never be so great as unity, for any value of F. Thus, for the curves derived by way of the method of limits,

$$W = \psi(F); \qquad \text{where} \qquad 0 < W \leqq 1,$$

for all values of the fundamental dimension of the stimulus (F) expressed in Fechnerian units, or Fechners.[9] These curves are, in this sense, graphical representations of the *Weber-Fechner function* for different sensory mechanisms.

Summary

1. Results of psychophysical research determining the Weber-ratio have been reviewed and described graphically for different sensory systems.

2. The Weber-ratio ($\Delta R/R$) is found to be a specific, dependent, and reproducible function of R in every particular instance. In all cases cited, as R increases, $\Delta R/R$ decreases and approaches a minimal value. Moreover, in the majority of these instances (there were but three exceptions), $\Delta R/R$ as a function or R tends to pass through this minimum and then to rise. For taste, the minimal value of $\Delta R/R$ has been specifically correlated with the introduction of a qualitative change. It is suggested that this correlation may obtain in other sense-departments.

3. Two important psychophysical factors are discussed: (*a*) the distinction between ΔR as mean j.n.d. and ΔR as variation-index, and (*b*) the need for an exact understanding of the relation which may exist between the two types of ΔR.

4. The stimulus to sensory intensity has been generalized for all sensory mechanisms in terms of the fundamental

[9] In designating the unit of the 'fundamental dimension' of the stimulus, the *Fechner*, we are following the suggested counsel of R. S. Woodworth (42, 10*f*).

dimension of the stimulus (F), the unit of which is designated the *Fechner*.

5. The general properties of W, the Weber-ratio that depends on ΔR as mean j.n.d., have been defined for any particular sensory system, and for all values of F.

REFERENCES

1. Aubert, H., Physiologie der Netzhaut, Breslau: Morgenstern, 1865, pp. xii+394.
2. Biedermann and Löwit, reported by Hering, E., Zur Lehre von der Beziehung zwischen Leib und Seele. I. Mittheilung. Über Fechner's psychophysisches Gesetz, *Sitzgsber. d. kaiserl. Akad. d. Wissen. Math. Natur. Classe*, 1875, **72**, Bd. 3, 310–348.
3. Blanchard, J., The brightness sensibility of the retina, *Phys. Rev.*, 1918, **11** (series 2), 81–92.
4. Boring, E. G., Did Fechner measure sensation?, Psychol. Rev., 1928, **25**, 242–251.
5. ——, The physical dimensions of consciousness, New York: Century, 1933, pp. xii+251.
6. ——, A history of experimental psychology, New York: Century, 1929, pp. xviii +699.
7. ——, Logic of the normal law of error in mental measurement, *Amer. J. Psychol.*, 1920, **31**, 1–33.
8. ——, Langfeld, H. S. and Weld, H. P., *et al.*, Psychology, a factual textbook, New York: Wiley, 1935, pp. xviii+455.
9. Crozier, W. J. and Hoagland, H., The study of living organisms, Chap. I, in 'A handbook of general experimental psychology' (ed. by C. Murchison), Worcester, Mass.: Clark Univ. Press, 1934, pp. 3–108.
10. ——, and Wolf, E., Dark adaptation in agriolimax, *J. Gen. Physiol.*, 1928, **12**, 83–109.
11. ——, On the variability of critical illumination for flicker fusion and intensity discrimination, *J. Gen. Physiol.* (in press).
12. Culler, E. A. K., Thermal discrimination and Weber's law, *Arch. Psychol.*, 1926, **13**, No. 81.
13. Fechner, G. T., Elemente der Psychophysik, Vol. II, Leipzig: Breitkopf and Härtel, 1860, pp. xii+571.
14. Fernberger, S. W., On the relation of the methods of just perceptible differences and constant stimuli, *Psychol. Monog.*, 1913, **14**, No. 4, 1–81.
15. Fletcher, H., Physical measurements of audition, *J. Franklin Inst.*, 1923, **196**, 289–326.
16. Gamble, E. A. McC., The applicability of Weber's law to smell, *Amer. J. Psychol.*, 1898, **10**, 82–142.
17. Gatti, A. and Dodge, R. Über die Unterschiedsempfindlichkeit bei Reizung eines einzelnen, isolierten Tastorgans, *Arch. f. d. ges. Psychol.*, 1929, **69**, 405–426.
18. Hecht, S., The visual discrimination of intensity and the Weber-Fechner law, *J. Gen. Physiol.*, 1924–25, **7**, 235–269.
19. ——, The nature of the photoreceptor process, Chap. 14, in 'A handbook of general experimental psychology' (ed. by C. Murchison), Worcester, Mass.: Clark Univ. Press, 1934, 704–828.

20. ——, and Wald, G., The visual acuity and intensity discrimination of drosophila, *J. Gen. Physiol.*, 1933–34, **17**, 517–547.
21. Holway, A. H. and Hurvich, L. M., Gustatory differential sensitivity to salt, *Amer. J. Psychol.*, 1936, **48** (in press).
22. Kellogg, W. N., An experimental comparison of psychophysical methods, *Arch. Psychol.*, 1929, No. 106, 1–86.
23. Knudsen, V. O., The sensibility of the ear to small differences of intensity and frequency, *Phys. Rev.*, 1923, **22** (2nd series), 84–102.
24. König, A. and Brodhun, E., Experimentelle Untersuchungen über die psychophysische Fundamentalformel in Bezug auf den Gesichtssinn, *Berlin Akad. d. Wissen. Sitzsber.*, 1888, **37** (2), 917–931.
25. ————, *ibid.*, 1889, **32** (2), 641–644.
26. Külpe, O., Outlines of psychology, New York: Macmillan, 1908, pp. xi+462.
27. Ladd, G. T. and Woodworth, R. S., Elements of physiological psychology, New York: Scribners, 1915, pp. xix+704.
28. Lemberger, F., Psychophysische Untersuchungen über den Geschmack, *Pflüg. Arch. f. d. ges. Physiol.*, 1908, **123**, 293–311.
29. Nutting, P. G., The complete form of Fechner's law, *Bull. Bureau of Standards*, 1907, **3**, 59–64.
30. ——, Luminous equivalent of radiation, *Bull. Bureau of Standards*, 1908, **5**, 261–308.
31. Lowry, E. M., The photometric sensibility of the eye and the precision of the photometric observations, *J. Opt. Soc. Amer.*, 1931, **21**, 132–136.
32. Oberlin, K. W., Sensitivity to lifted weights over a wide range of the stimulus, Harvard University Thesis, Harvard College Library, 1934, pp. 35.
33. Riesz, R. R., Differential intensity sensitivity of the ear for pure tones, *Phys. Rev.*, 1928, **31**, 867–875.
34. Skramlik, E. v., Handbuch der Physiologie der niederen Sinne, Leipzig: Thieme, 1926, pp. viii+532.
35. Stewart, G. W. and Hovda, O., Intensity factor in binaural localization, Psychol. Rev., 1918, **25**, 242–251.
36. Thomson, G. H., A comparison of psychophysical methods, *Brit. J. Psychol.*, 1912, **5**, 203–241.
37. Titchener, E. B., Experimental psychology: instructor's quantitative manual, New York: Macmillan, 1915, pp. clxxi+453.
38. Troland, L. T., Principles of psychophysiology, vol. 2, New York: Van Nostrand 1930, pp. xi+397.
39. Urban, F. M., Die psychophysischen Massmethoden als Grundlagen empirischen Messungen, *Arch. f. d. ges. Psychol.*, 1909, **15**, 261–355.
40. ——, *ibid.*, 1909, **16**, 168–227.
41. Weber, E. H., De pulsu, resorptione, auditu et tactu, *Annot. anat. et physiol.*, Leipzig, 1834.
42. Woodworth, R. S., A revision of imageless thought, Psychol. Rev., 1915, **22**, 1–27.
43. Zigler, M. J. and Holway, A. H., Differential sensitivity as determined by amount of olfactory substance, *J. Gen. Psychol.*, 1935, **12**, 372–382.
44. Zwaardemaker, H., Die Physiologie des Geruch, Leipzig: Engelmann, 1895, pp. vi+324.
45. ——, L'Odorat, Paris: Doin, 1925, pp. viii+305.

[MS. received September 21, 1935.]

THE MEASUREMENT OF DIFFICULTY

BY HERBERT WOODROW

University of Illinois

The psychologist who has observed a mental activity concerning which he would like further knowledge has only one procedure, namely, to vary the conditions of the activity and to note the effect thereon. The relation of the mental activity to the sum total of its conditions is the ultimate goal—sought, but never attained. The usual method, and at one stage, an indispensable one, is to vary only one condition while making the attempt (too often not completely successful) to hold all other conditions constant. Ultimately, however, these conditions must be summated. In nature several conditions vary simultaneously and in varying degree; and very soon therefore the problem arises of the interrelationship of all the conditions. To formulate this relationship in quantitative terms appears practically impossible unless the attributes of the observed mental activity are measurable in units which remain constant at all parts of the scale. For as soon as one condition is varied *in addition* to another, the second change occurs from a different initial point than the first; and not only is it not known whether the various changes will have algebraically summative effects, but it is impossible to determine the question unless the effect of the second change can be measured in comparable units at any two different levels —(1) that existing before the change in the first condition, and (2) that existing after such change. It is highly important, therefore, to have a scale uniform throughout for the measurement of the effect upon the attribute being studied of any changes superimposed upon any preëxisting constellation of conditions. As will be shown, simply using the standard deviation (or probable error) of the scores of a given group of subjects as the unit of measurement by no means leads to a constant unit scale. The probable error itself varies with variation in the conditions.

A large proportion of mental activities are what may be designated task-accomplishments. As illustrations may be mentioned the tasks of judging which of two stimuli is of greater magnitude, the task of reacting as quickly as possible to stimuli, all problem-solving, the tasks of memorizing, and all 'intelligence' tests. In all task-accomplishments one of the attributes of most importance is that which may be termed its goodness. Goodness may sometimes be identical with correctness, sometimes with speed, sometimes with the per cent of responses of a certain category, and so forth; but all task-performances are measurable in such a way that scores at one end of the scale constitute scores which are better, have more of the attribute 'goodness' than those at the other end. Now, in the case of task-performances, when the attribute under consideration is the goodness of the performance, then all the conditions which affect this attribute may be termed conditions of task-difficulty, that is the difficulty of performance of the task. The less the difficulty of the conditions, the better the score, or with reference to a fixed score, the greater the proportion of any group of subjects who will exceed that score. The difficulty of the conditions under which any task is performed, then, varies inversely with goodness of score (of a given subject or average score of a given group) and also inversely with the proportion of a population exceeding any fixed score in the task.

The present study deals with the problem of measuring the change in difficulty of a task-performance due to change in its conditions. It consists of a comparison of the results obtained by two independent methods of measuring the difference in difficulty of five different constellations of conditions. It is limited to the investigation of this matter in the case of one particular task, that of naming correctly a letter exposed tachistoscopically for .1 sec. The task chosen is one which recommended itself because of the considerable number of easily and readily varied conditions. Among these conditions are number of letters simultaneously exposed, amount of illumination, spatial separation, spatial pattern, and the duration of exposure. All of these conditions and their inter-

relationship have now been studied, but in the present investigation, dealing solely with the method of scaling difficulty, only the number of letters and their spatial separation have been varied. Through the combination of various numbers of letters and spatial separations, there were set up five different constellations of conditions of widely different difficulty.

The general conditions are described in some detail in a later paper and it will suffice for the present to state that instructions to the subject were to name the exposed letters, that the illumination was held constant throughout and that the bridge of the subject's nose rested against a support 70 cms. distant from the exposed letters. It should be stated that the sets of letters used under the various conditions were strictly comparable, each of the twenty letters of the alphabet which were employed occuring equally often in the case of every forty consecutive letters exposed. The letters themselves were black, Willson, gummed letters, ⅛-inch high, accurately placed on a dull white back-ground, along a horizontal axis. The five different sets of conditions, which will hereafter be referred to simply as *A*, *B*, *C*, *D* and *E*, were as follows:

	Number of Letters Simultaneously Exposed	Spatial Separation
A	2	2 cms.
B	2	3⅛ and 4 cms. (pooled)
C	4	4 cms.
D	4	10 and 12 cms. (pooled)
E	2	10 and 12 cms. (pooled)

Scores consisting of the percentage of letters correct under each of the above five conditions were obtained from 168 subjects, each instructed and tested individually. The total number of letters exposed in the case of each of the above sets of conditions (*i.e.*, the highest possible number correct) was 40 in all constellations except *D*, in which it was 80. Each subject was given preliminary practice consisting of 20 exposures, 10 of 2 letters and 10 of 4 letters. Each subject took the entire series of tests at one sitting which lasted approximately 60 minutes.

The raw data obtained are given in sufficient detail in

Table I, which shows the percentage of the group of 168 subjects making any of the obtainable scores when the latter are arranged in steps of 2½ per cent. Now everyone of the listed scores, with a few exceptions at the lowermost scores, is a score made by one or more subjects. In calculating the number of individuals exceeding any score, therefore, it has been necessary to assume that half the number of individuals actually making that score really exceeded it and that one-half failed to exceed it. This procedure is in accord with the hypothesis that if the scores were based on a very large number of trials without change of conditions and without practice effect, then half of the scores would be better than the obtained score (considered as precise to many decimals) and half worse. The application of this procedure results in a slight, desirable smoothing of the distribution of the group with respect to the obtained scores. All scores of 100 per cent correct, of which there were 45 with *A*, 4 with *B*, and 1 with *C*, are tabulated simply as exceeding 97½ per cent.

The problem, the answer to which is here sought, is the following: What is the difference in difficulty between the five constellations of conditions, *A*, *B*, *C*, *D* and *E*? The difficulty referred to is the difficulty of naming the letters correctly. The task is not considered as changing with the change in conditions; it remains that of naming correctly each and every letter presented. If the letters are presented in groups of four they will be more difficult to name; but that fact is here considered as a change in the difficulty of the conditions under which the performance of a constant task is attempted. Whether the methods to be described could be applied to a set of different tasks, *e.g.*, a pitch discrimination and a weight discrimination task, and if so, under what limitations, is a separate problem from the one here envisaged.

There are two methods which may be used to measure the difference in the median difficulty, or the difficulty for the median subject, of the above task. For convenience these methods will be termed the population percentage scaling method and the score percentage scaling method. It will be shown that the two methods, based on separate assumptions, lead to approximately the same values.

TABLE I

VARIATION WITH CONDITIONS IN THE PERCENTAGE OF THE GROUP EXCEEDING THE GIVEN SCORES

Score (Per cent correct)	Constellation of Conditions				
	A	*B*	*C*	*D*	*E*
97.5	35.7	6.8	0.9		
95.0	51.5	18.5	2.4		
92.5	64.6	29.2	5.1		
90.0	72.9	39.3	8.9		
87.5	78.6	50.0	13.1		
85.0	84.2	57.7	17.6		
82.5	87.5	65.2	22.6		.0
80.0	89.6	72.0	28.6	.0	.6
77.5	91.4	76.2	34.5	1.2	1.2
75.0	93.5	78.9	39.6	1.5	1.5
72.5	94.9	82.4	44.6	3.0	3.0
70.0	95.8	85.1	51.5	5.4	5.1
67.5	97.0	86.9	58.9	8.6	8.0
65.0	98.2	89.3	64.9	13.7	12.8
62.5	99.1	91.1	70.5	18.5	17.3
60.0	99.7	92.3	75.0	22.9	21.7
57.5	100.0	93.5	78.9	30.1	26.2
55.0		94.3	83.0	38.4	31.0
52.5		95.2	86.3	46.7	37.5
50.0		96.4	88.7	54.2	44.3
47.5		97.3	90.8	61.3	50.6
45.0		97.6	93.2	68.5	58.6
42.5		97.9	95.2	73.5	68.5
40.0		99.1	97.3	78.3	75.6
37.5		100.0	99.1	82.7	80.1
35.0			99.4	87.8	84.8
32.5			99.7	91.1	89.0
30.0			100.0	93.5	92.0
27.5				96.4	95.2
25.0				97.0	97.3
22.5				98.8	98.2
20.0				100.0	99.1
17.5					99.7
15.0					100.0

The assumption on which the population percentage scaling is based is simply the very familiar one that the distribution of ability in the group used for the purpose of scaling is normal. By definition the difficulty of making any score varies directly with the amount of ability required to make it. The abilities represented by various scores, on the assumption of a normal distribution, vary inversely with the proportions of the population exceeding those scores as points on the base-line of a normal distribution with reference to the propor-

tion of the area lying above (or to the right) of them. Thus, if a score is exceeded by only 25.0 per cent of the population it represents an ability of + 1 P.E. of the distribution of ability, and if exceeded by 75 per cent of the population, the individual making it has an ability of − 1 P.E. But since the difficulty of a score, by definition, varies directly as the ability required to make it, the distribution of difficulty must be identical with the distribution of ability. A score made by 75 per cent of the group, therefore, has a difficulty of − 1 P.E. of the distribution of difficulty of the population of individual scores. Ability is reflected only in goodness of score; and to say that ability is normally distributed is the same as saying that the scores made by the individuals are normally distributed as regards their goodness; and this in turn is the same as saying that the scores are normally distributed as regards the difficulty of making them. It is in this last form that the hypothesis affords a basis for measuring difficulty. The concept utilized is that of a normal distribution of a population of scores when these scores are spaced along a scale graduated in units of difficulty. For the present purpose, then, the familiar assumption that ability is normally distributed becomes the less familiar one that the scores made by the various individuals composing the group are normally distributed as regards their difficulty. Stated in this latter way, the assumption permits, in place of the scaling of ability, the scaling of difficulty; and in place of the measurement of the difference in mean ability between two groups of subjects, the measurement of the change in difficulty represented by a change in the mean score of a single group due to change in conditions. In the case of any one set of conditions, the difficulty of making any score may be readily obtained on the basis of the above stated assumption of a normal distribution. All that is necessary is to place the scores on the base-line of a normal distribution area so that the area to the right equals the proportion of the population passing that score, and the area to the left the proportion failing. The proper values are obtained from the Kelley-Wood tables, showing the distances to the right or left of the

mean (x-values) required in order that given proportions of the normal area shall lie above or below these points.

This scaling of the difficulty of the various *scores*, obtained under each of five sets of conditions, constitutes the first step to be taken in the case of the present data. In Table II each of the population percentages exceeding each score has been transformed into a standard deviation score on the above assumption. These values will hereafter be termed the population x-values. The values for each constellation of conditions is in terms of the standard deviation in difficulty of

Table II

Standard Deviation Values of the Percentages of the Group Exceeding the Given Scores

Score (Per cent correct)	Constellation of Conditions				
	A	*B*	*C*	*D*	*E*
97.5	− .366	−1.491	−2.366		
95.0	+ .038	− .896	−1.977		
92.5	+ .375	− .548	−1.635		
90.0	+ .610	− .272	−1.347		
87.5	+ .793	.000	−1.122		
85.0	+1.003	+ .194	− .931		
82.5	+1.150	+ .391	− .752		− ∞
80.0	+1.259	+ .583	− .565	− ∞	−2.512
77.5	+1.366	+ .713	− .399	−2.257	−2.257
75.0	+1.514	+ .803	− .264	−2.170	−2.170
72.5	+1.635	+ .931	− .136	−1.881	−1.881
70.0	+1.728	+1.041	+ .038	−1.607	−1.635
67.5	+1.881	+1.122	+ .225	−1.366	−1.405
65.0	+2.097	+1.243	+ .382	−1.094	−1.136
62.5	+2.366	+1.347	+ .539	− .896	− .942
60.0	+2.748	+1.426	+ .674	− .742	− .782
57.5	+ ∞	+1.514	+ .803	− .522	− .637
55.0		+1.580	+ .954	− .295	− .496
52.5		+1.665	+1.094	− .083	− .319
50.0		+1.799	+1.211	+ .105	− .143
47.5		+1.927	+1.329	+ .287	+ .015
45.0		+1.977	+1.491	+ .482	+ .217
42.5		+2.034	+1.665	+ .628	+ .482
40.0		+2.368	+1.927	+ .782	+ .693
37.5		+ ∞	+2.366	+ .942	+ .845
35.0			+2.512	+1.165	+1.028
32.5			+2.748	+1.347	+1.227
30.0			+ ∞	+1.514	+1.405
27.5				+1.799	+1.665
25.0				+1.881	+1.927
22.5				+2.257	+2.097
20.0				+ ∞	+2.366
17.5					+2.748
15.0					+ ∞

the 168 scores made under those particular conditions. There is no justification for the assumption that this standard deviation does not vary with change in conditions. In fact, a careful study of the data of Table II indicates just such a variation.

Before proceeding further with the determination of the difference in difficulty due to change in conditions, it is desirable to examine the relation between the difficulty values of the various scores made under different conditions. It is obvious from inspection that a high correlation exists between these values for the fixed set of scores given in the left-most column, headed 'score,' which includes all the obtained scores by steps of 2½ per cent from 97½ downwards. The coefficients of correlation between the x-values of adjacent columns are as follows: $r_{AB} = .978$; $r_{BC} = .990$; $r_{CD} = .984$; and $r_{DE} = .998$. If the x-values for all scores exceeded by more than 92½ per cent or less than 7½ per cent of the population be excluded, that is, all beyond the range $\pm 1.44\sigma$,

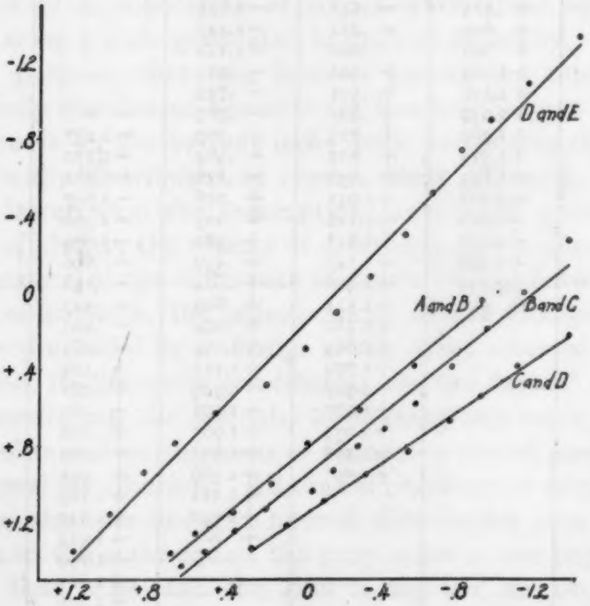


FIG. 1. The relation between the x-values of a fixed set of scores under conditions A and B, B and C, C and D and D and E.

the correlation between the x-values of adjacent columns is in all cases over $+.99$. The relation between the x-values, within this range, for adjacent columns is shown by Fig. 1 closely to approximate linearity.

In this figure, ordinates represent the x-values for the set of conditions nearest A and the abscissæ those for the set nearest E. For example, in the case of the curve labelled 'B and C,' the x-values for B are the ordinates and those for C the abscissæ, while in the curve labelled 'C and D,' the x-values for C are the ordinates and those for D the abscissæ; etc. Since the correlation between the plotted coördinate values is practically $+ 1$, the line of relation, or regression, may be expressed without appreciable error by the formula

$$x_A = x_B(s_A/s_B)$$

in which x_A and x_B are the x-values for any given score and s_A and s_B the values $\sqrt{\Sigma d^2/n}$, calculated from the mean of the x-values of A and B, respectively. In Fig. 2, each set of points for paired coördinates has been fitted with a straight line following this formula, a line which passes through the point representing the intersection of the means of the two sets of x-values.

The results obtained up to this point could be represented by a diagram, on which the base-line would indicate the difficulty of the various percentage scores, and the frequency curves for each constellation of conditions would be drawn so that that portion of any curve to the right of any specified score would indicate the percentage of all the scores made under that set of conditions which exceeded the specified score. If the diagram were constructed from the data of Table II, a difficulty would arise due to the fact that the correlation of the difficulty values of a given set of scores under two different sets of conditions is not perfect. This lack of perfect correlation means that slightly discrepant scales would have to be used for representing the difficulty of the various scores in the case of the different sets of conditions. These discrepancies are small, however, and the final calculated values will permit the drawing of curves on the same

base-line in such a way that the discrepancy between the obtained values and the calculated ones represented by the curves may be ignored as probably within the range of deviation between the assumed perfect normality of distribution under each of the five conditions of difficulty and the actually prevailing distribution.

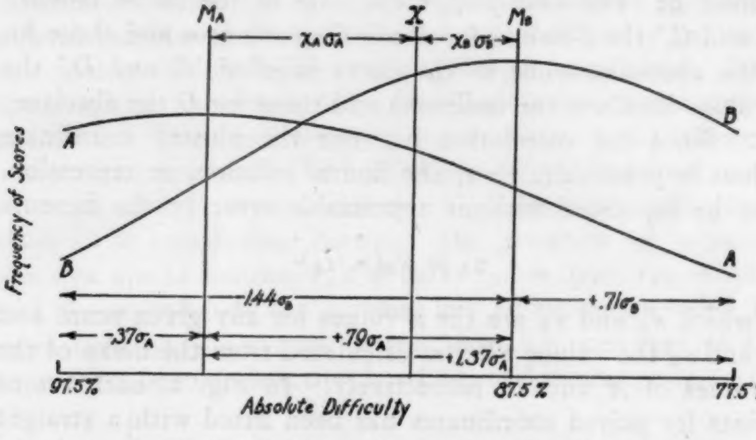


FIG. 2. Absolute difficulty of making certain scores under two different constellations of conditions, *A* and *B*.

Figure 2 illustrates the situation. For the sake of simplicity, it is limited to the data obtained under conditions *A* and *B*, and the three scores, 97.5, 87.5 and 77.5 per cent, correct. The diagram shows that it is possible to represent to a very close approximation all the data contained in Table II for these three scores under the two conditions of difficulty *A* and *B* by curves drawn on a single base-line. The single discrepancy between the graphic values and the values of Table II is that in the graph, the distance on the base-line from score 87.5 to 97.5 indicates an x-value in the case of *B* of $-$ 1.44; whereas to conform with the data of Table II it should indicate a value of $-$ 1.49. This slight discrepancy may be ignored. The base-line of this diagram represents absolute difficulty. That per cent scores may be placed along it means that as the per cent scores fall, difficulty rises, so that each different score represents a different degree of difficulty and also that the correlation between the difficulty of

a given set of scores in A and B is approximately $+1.0$. The curves A and B, as drawn, are portions of normal distribution curves. Each curve represents the distribution of difficulty of the obtained 168 scores on the assumption that this distribution is normal (or what amounts to the same thing, on the assumption that the distribution of ability of the group of subjects producing these scores is normal). The standard deviation of any one curve, then, is the standard deviation in the difficulty of making the 168 scores under one particular set of conditions. The greater height of curve B at the median is dictated by the fact that the standard deviation of the B curve is less than that of the A curve ($y_0 = .399N \cdot \sigma_A/\sigma_B$).

The point on the base-line of Fig. 2 marked X represents the difficulty of any score. Obviously, then, $X = M_A + x_A\sigma_A$; that is to say, the point X lies above the median of curve A, represented by M_A, by x_A times the σ_A, in which x_A is the population x-value of the percentage score having the difficulty X. Likewise, $X = M_B + x_B\sigma_B$ (x_B would be negative for the particular X shown in Fig. 2). It follows, since any X is a fixed position, that

$$M_A + x_A\sigma_A = M_B + x_B\sigma_B$$

or, transposing,

$$M_B - M_A = x_A\sigma_A - x_B\sigma_B \quad . \quad . \quad . \quad (1)$$

Since this equation holds of any X value, the value $M_B - M_A$ should theoretically be the same for x-values calculated for any score (*cf.* Table III). The equation may be solved for all the scores made under both A and B, and the resulting values averaged for a final determination of the difference in difficulty between A and B, *i.e.*, $M_B - M_A$. Therefore,

$$M_B - M_A = \left(\frac{\Sigma x_A}{n}\sigma_A - \frac{\Sigma x_B}{n}\sigma_B\right)$$

or

$$M_B - M_A = \left(\frac{\Sigma x_A}{n} - \frac{\Sigma x_B}{n} \cdot \frac{\sigma_B}{\sigma_A}\right)\sigma_A \quad . \quad . \quad . \quad (2)$$

The values $\frac{\Sigma x_A}{n}$ and $\frac{\Sigma x_B}{n}$ are the means of the x-values of A and B, and may be represented by m_A and m_B. And the value, $\frac{\sigma_B}{\sigma_A}$, may be represented in terms of the x-values. The variation in the x-values of a given set of points on the scale of difficulty will of course vary inversely with the dispersion of the group over that scale, since the greater the dispersion of the group, the smaller the proportion of the group which will be subtended by any two points, and therefore the less the difference in the x-values of those two points. The x-values for a given set of scores under any one set of conditions will, then, vary inversely with the standard deviation of the group with respect to the scale on which the difficulty of these scores is located. In the case of two different distributions, for sets of x-values between which the correlation is $+1.0$ as may be assumed without appreciable error to be the case in the present instance, as Thurstone has shown[1] the following equation may be written:

$$\frac{\sigma_B}{\sigma_A} = \frac{s_A}{s_B}$$

in which

$$s_A = \sqrt{\frac{\Sigma d_A^2}{n}} \quad \text{and} \quad s_B = \sqrt{\frac{\Sigma d_B^2}{n}}$$

and d is the deviation of the x-values from their mean. Substituting m_A for $\frac{\Sigma x_A}{n}$, m_B for $\frac{\Sigma x_B}{n}$ and $\frac{s_A}{s_B}$ for $\frac{\sigma_B}{\sigma_A}$, in equation (2), we arrive at the difference in difficulty between any two sets of conditions, A and B.

$$M_B - M_A = \left(m_A - \frac{s_A}{s_B} m_B \right) \sigma_A \quad . \quad . \quad . \quad (3)$$

Equation (3) is, and has intentionally been made to be, exactly the equation devised by Thurstone for measuring the difference in mean ability between two groups of subjects,

[1] L. L. Thurstone, A method of scaling psychological and educational tests, *J. Educ. Psychol.*, 1925, 16, 440.

such as adjacent age groups. If the data had been obtained from two different groups of subjects, *A* and *B*, tested under identical conditions, then, in fact, the above equation applied to the data would give the difference in mean ability between the two groups. Actually, since only one group has been employed, but tested under the two sets of conditions *A* and *B*, the difference between *A* and *B* is not properly to be regarded as due to a change in the ability of the group but as due to the change in the difficulty of conditions. After the increase in difficulty of *B* over *A* has been determined, then M_C may be substituted for M_B and M_B for M_A; and so on. As the value of M_A, the median difficulty of constellation *A*, any value may be assumed since absolute scaling does not give values measured from an absolute zero but from an arbitrary starting point, and it lays off distances from this arbitrary starting point merely in equal or constant units, the magnitude of which is arbitrary. According to the above formula, the difference in difficulty between *B* and *A* is to be measured in the arbitrary unit consisting of the standard deviation in difficulty of the scores obtained under the set of conditions represented by *A*.

In obtaining the difference in the mean of the *x*-values under any two sets of conditions, corresponding *x*-values must of course be used, that is to say, *x*-values for but one set of scores. Moreover, it is undesirable to use the *x*-values for scores lying near the extreme ends of the distribution on account of the relatively great unreliability of the frequencies of such scores. In the present study, the scores which have been used include all the obtained scores, by steps of 2½ per cent, which were not exceeded by more than 92½ per cent or less than 7½ per cent of the group. In other words those scores were used which had *x*-values not exceeding (in the case of either of the compared distributions) the value 1.44. The available *x*-values for determining the difference in difficulty between *A* and *B* were eight in number. Their mean in the case of *A* is + .8243 and in the case of *B*, + .0206. The ratio of $\frac{s_A}{s_B}$, s_A applying to the used *x*-values of *A* and s_B to

those of B, is $\frac{.431}{.526}$, or .8194. It remains to multiply the mean of the B x-values by this ratio and subtract it from the mean of the A x-values, *i.e.*, to calculate the value of $\left(m_B - \frac{s_A}{s_B} m_A\right)$. The result is $\left(.8243 - \frac{.431}{.526} \times .0206\right)$ or .8074. Thus, the conclusion is reached that in units of the standard deviation of A, B is .8074 more difficult than A.

At this point a slight difficulty arises in that it is impossible directly to compare the x-values of A with those for the more difficult conditions, since the difference in difficulty between A and either D or E is so great that there is no overlap in the distributions except at the unreliable tails. It is necessary, therefore, to use a dovetail method: instead of directly determining how much more difficult C is than A, to determine first how much more difficult C is than B in terms of σ_A, then how much more difficult D is than C, etc. For determining the difference in difficulty between B and C, the 13 x-values for the percentage scores 90 to 60, inclusive, are available; for comparison of C and D the 9 x-values for the percentage scores 67.5 to 47.5, inclusive; and for comparison of D with E, the 15 x-values for the scores 67.5 to 32.5 inclusive. The results of the various steps in the computations are as follows:

$$\sigma_B/\sigma_A = s_A/s_B = .8194,$$

$$\sigma_C/\sigma_B = \frac{.510}{.619} = .8239, \quad \sigma_C/\sigma_A = .8194 \times .8239 = .6751,$$

$$\sigma_D/\sigma_C = \frac{.357}{.527} = .6749, \quad \sigma_D/\sigma_A = .6751 \times .6749 = .4556,$$

$$\sigma_E/\sigma_D = \frac{.822}{.794} = 1.035, \quad \sigma_E/\sigma_A = .4556 \times 1.035 = .4715.$$

To calculate the difference in difficulty between C and B, the means of the corresponding x-values of each are reduced to units of the standard deviation of the distribution in the case of A. Thus,

$$\frac{M_C - M_B}{\sigma_A} = m_b \frac{s_A}{s_B} - m_c \frac{s_A}{s_B} \cdot \frac{s_B}{s_C}$$
$$= (+ .7325 \times .8194) - (- .2814 \times .6751) = .7902.$$

The final outcome of the calculations is a list of the amounts by which the difficulty of each set of conditions exceeds the difficulty of the next less difficult set, all the differences being expressed in terms of the $\sigma_{dis.}$ of the difficulty of the scores obtained under the constellation of conditions *A*. The results may be expressed as follows:

Change in Conditions	Change in Absolute Difficulty
From *A* to *B*	.807
From *B* to *C*	.790
From *C* to *D*	.774
From *D* to *E*	.065

Since the order of difficulty is from *A* to *E*, the magnitude of other steps in difficulty, such as those from *A* to *C*, *B* to *D*, *A* to *D*, etc., may be readily obtained. For example, the increase in difficulty produced by change from *A* to *C* is 1.597 and that from *A* to *E* is 2.436. In short, the difficulty of the task under any of the conditions *A* to *E* may now be located at points on a straight line representing a constant unit scale of difficulty.

It will be observed that the above method of measuring difficulty makes no assumption concerning the difference in absolute difficulty which is represented by any two percentage right scores. How much difference in difficulty corresponds to any given difference in the scores has to be determined by scaling; it is not assumed.

One of the most interesting results of Thurstone's scaling method was the proof, incidental to the actual scaling, that absolute variability in ability increases with age (the percentage, or relative, variability tending to remain constant). Thurstone makes use of successive age groups. He takes as a fixed step in ability the difference in ability required to pass two different tests (or to make two different scores on the same test under *one set of conditions*). He then shows, in substance, that the variability of an older group is greater than that of a younger group because this fixed step in difficulty subtends a smaller proportion of the older group. In the present study, on the other hand, with a constant group of subjects, it has been found that the percentage of the group

Table III

Absolute Difference in Difficulty of Obtaining Various Scores in Terms of the Standard Deviation of Difficulty under Constellation of Conditions *A*

Score (Per cent correct	Difficulty Step			
	A to *B*	*B* to *C*	*C* to *D*	*D* to *E*
97.5				
95.0	+.772			
92.5	+.824			
90.0	+.833	+.686		
87.5	+.793	+.757		
85.0	+.844	+.788		
82.5	+.830	+.828		
80.0	+.781	+.859		
77.5	+.782	+.853		
75.0		+.836		
72.5		+.855		
70.0		+.827		
67.5		+.767	+.774	+.040
65.0		+.761	+.756	+.038
62.5		+.740	+.772	+.036
60.0		+.713	+.793	+.031
57.5			+.780	+.062
55.0			+.778	+.100
52.5			+.777	+.112
50.0			+.770	+.115
47.5			+.766	+.124
45.0				+.118
42.5				+.059
40.0				+.029
37.5				+.031
35.0				+.046
32.5				+.035
Mean	+.807	+.790	+.774	+.065
M.V.	.025	.055	.007	.032

subtended by any given difference in score varies with the conditions of the task-performance. At first thought, the present data might appear to justify the generalization that the absolute variability in task-performance decreases as the task becomes more difficult. The data are not suited, however, to establish final conclusions on this matter. It is quite possible that change in the number of letters from 2 to 4 has some effect upon variability which is independent of its effect upon difficulty. As will be pointed out in a later study,[2] there is a greater chance of variation in the distribution of

[2] The interrelation of conditions of difficulty: II. number, spatial separation and illumination as conditions of simultaneous letter span, *J. Gen. Psychol.* (in press).

attention in the case of 4 than in that of 2 letters. Variability then is probably not solely a function of difficulty. It should be pointed out, moreover, that there is no obvious reason why the relation between variability and difficulty should be the same for all measurable attributes of performance. The attribute measured in the present study—correctness—is one which has finite limits of zero and 100 per cent. Other attributes—for example, spatial or temporal attributes, such as distance and time—do not appear to be so limited; and it would be rather surprising if this fact did not affect the relation between variability and difficulty. However this may be, the important fact that variability varies enormously with the conditions of difficulty of the task appears to be established beyond question by the preceding data.

The fact of change in the variability of a given group with change in the conditions of difficulty is undoubtedly related to Thurstone's generalization that variability increases with age. For it is obvious that for older children tests are less difficult than for younger children, since they are 'passed' by a larger percentage of the older children. In this case, then, variability of the group depends upon the relation of the median ability of the group to the degree of ability required to pass the tests, that is to say, their difficulty. The greater this ratio, the greater the variability.

The fact that a given set of tests is less difficult for older than for younger children suggests an alternative formulation of Thurstone's law of increase in invariability with age. Just as one might consider a test taken after practice to be taken under changed conditions from the same test taken before practice, so one might consider tests taken by older children to be taken under different (subjective) conditions, than when taken by younger children, and to consider that the two tests, having changed as regards their conditions, may not be regarded as necessarily constituting a fixed step in difficulty. Difficulty does not attach to tests considered as situations outside of a subject but to the performance of the tests; and one important set of conditions of goodness of performance consists of conditions within the subject (which may be termed

subjective conditions). That, when these conditions vary, the size of the step in difficulty between two tests may change, is just as certain as that it may change with change in objective conditions. From this point of view, an alternative way of expressing Thurstone's generalization that variability increases with age would be to say that variability in goodness of performance (ability) increases as performance becomes less difficult (because of change in those subjective conditions of difficulty of performance which are affected by age). In this latter form, the generalization would appear to fit very well the increase in variability of a given group with practice (*i.e.*, with decrease in difficulty due to practice).

The primary interest of the present study is not in the relation of variability to difficulty, but in the measurement of difficulty. As stated earlier, the chief purpose is to compare the absolute difficulty values obtained by the population percentage scaling method, which have been given above, with those obtained by the score percentage method, which have yet to be indicated.

The score percentage scaling method, as here conceived, ignores the individuals composing the group and considers solely the responses. Each letter named, correctly or incorrectly, is a response. The statistical population now considered is not the population of individuals but the population of responses. The theory of the method remains the same, whether these responses are those of a single individual, of a more or less homogeneous small group of individuals or the responses of a large and more or less normally distributed group of individuals as in the present instance. It may be most readily stated by assuming that the responses referred to are those of a single individual, though ordinarily this individual will be an 'ideal' individual, one corresponding to the mean or the median of the group of individuals used as subjects. Results from such an ideal or average individual will be more uniform and less subjected to inexplicable variations from sitting to sitting than those obtained from any one concrete individual. In the present study the measurements which have been made are those of difficulty for the

median individual, a hypothetical individual whose performances under all conditions are identical with those of the median of the group of 168 subjects.

The theory is simply that the population of responses is normally distributed, as regards the amount of ability represented by each response. The greater the ability represented by a response the greater may be the difficulty of the conditions under which a letter may be presented and yet be named correctly. The percentage correct will then agree exactly with the percentage of responses representing an ability equal to or in excess of the amount of ability required to name the letter correctly. As the difficulty of the conditions is increased, the percentage of the responses able to cope with or surmount the existing difficulty will decrease, so that degree of difficulty varies inversely with the percentage correct. Since the amounts of ability represented by the separate responses are assumed to be normally distributed, the percentage of correct responses made by one hypothetical individual under any set of conditions equals the portion of a normal distribution area lying above (or to the right of) a point on the base-line representing the amount of ability required to make a correct response, or, what is the same thing, a point representing the degree of difficulty of the letters giving that percentage correct. The scaling of difficulty produced by any set of conditions can, then, on this hypothesis, be accomplished by setting the difficulty at such a point on the base-line of a normal distribution that the percentage correct (or the mean or median percentage correct in the case of a group of subjects) corresponds to the portion of the distribution area lying to the right of that point. The distance of the required point from the mean of the distribution (an arbitrary X representing the degree of difficulty yielding 50 per cent correct responses) may be read directly from the Kelley-Wood tables.

Since in the present investigation several of the conditions gave scores on the part of some subjects as high as 100 per cent correct and the distribution of the per cent scores in these cases is skewed to a rather extreme degree, it is necessary to use the

median per cent correct in place of the mean. Even the medians can be determined only approximately, since the scores are really discrete scores with many ties at the scores near the median. Interpolation is required. It may be done in many ways, but the method here used, while probably not the most reliable, is the simplest: namely, simple linear interpolation between the two scores between which the median lies. The median percentages correct, and the corresponding x-values, as obtained by this method of interpolation, from the data presented in Table I, are as follows:

Conditions	Median Per Cent Correct	Absolute Difficulty (Score per cent x-value)
A	95.24	−1.669
B	87.50	−1.150
C	70.54	− .540
D	51.40	− .035
E	47.74	+ .057

In order to compare these results by the score percentage method with those obtained by the population percentage method, it is necessary to express them in terms of the same unit. The obtained difference in difficulty between *A* and *E* is a smaller number of units by the score per cent method than by the population per cent method, being 1.726 by the former and 2.436 by the latter. But both the unit and the zero point are arbitrary. All that is really measured is the relative magnitude of the distances between any two of the difficulties represented by *A*, *B*, *C*, *D* and *E*. These distances can of

TABLE IV

RESULTS OF TWO METHODS OF MEASURING DIFFICULTY EXPRESSED IN TERMS OF EQUIVALENT UNITS

Difficulty Step	By Score Per Cent Scaling	By Population Per Cent Scaling	Discrepancy
From *A* to *B*	.515	.572	+.057
From *B* to *C*	.614	.559	−.055
From *C* to *D*	.505	.549	+.044
From *D* to *E*	.092	.046	−.046
Total range, *A* to *E*	1.726	1.726	
From *A* to *C*	1.129	1.131	+.002
From *A* to *D*	1.634	1.680	+.046
From *B* to *D*	1.119	1.108	−.011
From *B* to *E*	1.211	1.154	−.057
From *C* to *E*	.597	.595	−.002

course readily be reduced to the same scale. If this is done by multiplying those obtained by the population percentage method by 1.726/2.436 or .7085, then the results obtained by the two methods correspond as indicated in Table IV. The closeness of the correspondence is clearly indicated by Fig. 3.

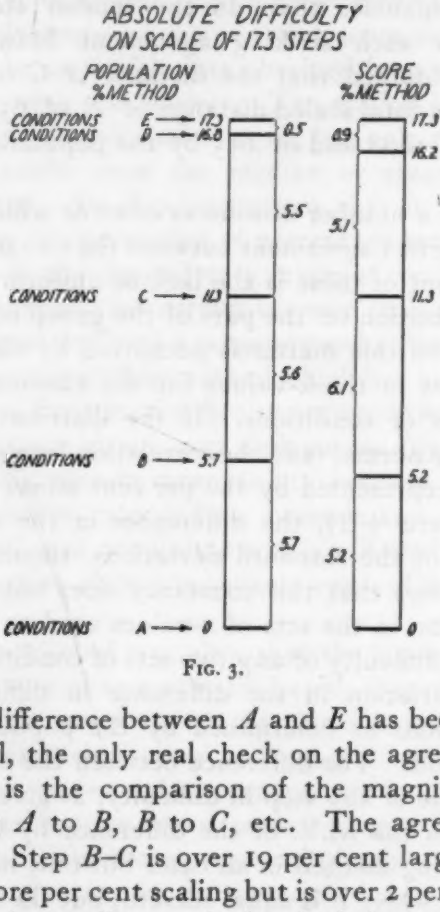


FIG. 3.

Since the difference between *A* and *E* has been arbitrarily made identical, the only real check on the agreement of the two methods is the comparison of the magnitudes of the separate steps *A* to *B*, *B* to *C*, etc. The agreement is far from perfect. Step *B–C* is over 19 per cent larger than step *A–B* by the score per cent scaling but is over 2 per cent smaller by the population scaling method. And the very small step *D–E* is twice as large by the one method as by the other. Nevertheless, the maximum discrepancy of .057 is but 3.30 per cent of the entire distance scaled. Since this distance is less than one-half of the scalable range with a given group of subjects, it may be said that if the whole scalable range in difficulty were divided into 100 steps, the position of any given

constellation of conditions of difficulty on this scale would not be affected by more than a maximum of 1.65 steps by the method used in determining it. The obtained discrepancies fluctuate in direction, so that wide differences in difficulty measured by the two methods do not, on the average, show greater discrepancies than do the smaller steps. Thus it happens that such striking agreement between the two methods is obtained that the distance *A–C* constitutes the fraction of the total scaled distance, *A–E*, of .654 by the score percentage method and of .655 by the population percentage method.

There are a number of sources of error which account for the lack of perfect agreement between the two methods. The most important of these is the lack of uniform and perfectly normal distribution on the part of the group of 168 subjects. A judgment on this matter is permitted by the variation in the differences in the x-values for the various scores under any two sets of conditions. If the distribution of ability were perfectly normal (and the correlation between the degrees of difficulty represented by the per cent scores under various conditions were $+1$), the differences in the x-values, after equalization of the standard deviations, should be constant. Table III shows that this constancy does not prevail. The mean variation in the sets of x-values used to determine the difference in difficulty of any two sets of conditions represents the mean variation in the difference in difficulty between those conditions as determined by the population per cent scaling method. The difference between the two methods in the magnitude of any step in difficulty, as given in Table IV, is larger than the M.V. of the difference by the population per cent scaling method in all cases but one, namely the step from *B* to *C*, where it is equal thereto; but the tails of the distributions have been omitted and errors in the score per cent method due to using interpolated medians also play a part in accounting for the lack of complete agreement between the two methods.

Probably the chief difficulty in the population per cent scaling method is that to use it one must first obtain scores

from a very large number of subjects. There always remains the risk of obtaining a group which does not satisfactorily approximate a normal distribution. Its advantage lies in the fact that it may be used with any type of scores, or with tests which are simply passed or failed. Also, by using different age groups, it may be used over a very wide range of conditions. The advantage of the per cent score scaling method is that it may be used with data obtained from a small group of subjects, or even one subject. The scaling is exceedingly simple, requiring no calculation at all, but only the use of the Kelley-Wood tables once the median or mean percentage right is obtained. Its disadvantage is that it is limited, in this direct form, to the scaling of percentage scores. With a given group it is also limited to a range of conditions which allow median percentages correct of more than zero and less than 100, and practically to a range somewhat less. However the score percentage scaling method which has here been used to measure the difference of difficulty between two sets of conditions for a given group may just as readily be used to measure the difference in mean ability between two groups. It may therefore be extended to measure any difference in degree of difficulty of conditions which will enable any individual of any age or ability to make a score of over zero and less than 100.

It should be pointed out, also, that the concept on which the score per cent scaling method is based, that of the normal distribution of the separate responses, leads to a third method of absolute scaling which may be used whenever a value is given to each separate trial or response, as in shooting at a target or in reaction-time experiments. It may be used with data from a small homogeneous group of subjects or even a single subject, and either for the purpose of measuring difference in difficulty of conditions or difference in mean ability of individuals or groups of individuals. The method consists in using formula (3) as given above, but the x-values are determined, not from the percentages of a large population of subjects making the various scores, but from the percentages of the pooled responses exceeding a certain value. This third

method, which may prove to have a wider range of convenient application than either of the two methods used in the present study, is now being applied to the study of the interrelation of conditions of difficulty in the simple reaction-time experiment.

Agreement between two methods does not prove the validity of either but it does add to the plausibility of each since it shows that each method harmonizes with a plausible assumption other than the one on which it is based. The population per cent scaling method is based on the assumption of a normal distribution of the mean ability of individuals; the score per cent scaling method is based on the assumption of a normal distribution of the ability at successive trials of a given individual, or group of individuals. While the agreement in outcome indicates that these two assumptions in the present instance are in harmony with each other, it does not follow that they are identical. They appear to be separate assumptions, and if so, it would be impossible for the methods to agree if one were correct and the other incorrect; and if both were incorrect the agreement in results would be a very improbable coincidence.

It remains to emphasize one point by way of caution. The score per cent scaling method as here used, measures the difficulty of a task, for a single individual, or for that 'ideal' individual corresponding to the mean or the median of any group of individuals. With a given group of subjects, it enables one to say, starting with any desired set of conditions, just how much any change in the conditions changes the difficulty of the task. It does not, however, tell anything about the difference in ability of two individuals making different scores under a given set of conditions. It therefore does not indicate that two individuals making scores, say of 60 and 50 per cent, under one set of conditions differ in ability the same amount as two individuals making 60 and 50 per cent scores under a different set of conditions. All it does is to give a measure of difference in difficulty for the mean, or as applied in the present study, the median, individual.

The score per cent scaling method described in this paper

has been used in a series of studies [3] of the effect of combined changes in a number of conditions. The problem attacked has been the very general one of the relationship between the goodness of a performance and the sum total of its conditions. A number of conclusions have been reached. One of these is that an increase in difficulty due to a fixed change in any one condition will increase the total difficulty of performance by the same amount, no matter what the amount of difficulty due to the combined remaining conditions. Without a method of measuring absolute difficulty, the attempt to study this sort of problem experimentally would be to encounter a hopelessly bewildering maze of incomparable changes in raw percentage scores.

Summary

Through variations in number and spatial separation, five widely different degrees of difficulty were given to the task of naming letters presented tachistoscopically for .1 sec. Scores were obtained from 168 university students tested individually under each of the five sets of conditions. The difficulty of these five sets of conditions was then determined by two quite different methods. One method, based on Thurstone's procedure for calculating the difference in mean ability between two groups of subjects, yields a measure of difference in terms of the standard deviation of the difficulty of the task-performance under one set of conditions. The other method is one which assumes that the relation between the average per cent correct score and the difficulty of the conditions may be represented by the ogive curve of a normal distribution, of which the ordinates indicate the average percentage correct scores and the abscissæ indicate the inverse of the difficulty of the conditions resulting in those average scores. The two methods were found to yield results in substantial agreement.

In addition, a third method usable under certain conditions has been described. All three methods may be applied either to the scaling of difficulty or the scaling of ability.

[MS. received October 14, 1935]

[3] The interrelationship of conditions of difficulty: I. The effect of change in number at various spatial separations in simultaneous letter span; and II. Number, spatial separation and illumination as conditions of simultaneous letter span, *J. Gen. Psychol.* (in press). Other studies will follow.

THE NATURE OF THE CONDITIONED RESPONSE: I. THE CASE FOR AND AGAINST STIMULUS-SUBSTITUTION

BY ERNEST R. HILGARD

Stanford University

In order to achieve clarity in the discussion of theories of learning, it is necessary to make some sort of preliminary classification. According to Tolman (26), the theories fall into two groups, reflex-arc and field theories, the former being subdivided into conditioned reflex and trial-and-error types. The field theories are essentially Gestalt-like. Dashiell (6) has similarly classified theories as conditioned response, trial-and-error and Gestalt. This study is directed to the examination and interpretation of one of these contending theories, that of conditioned response, in the light of experimentally revealed data.

Tolman proceeds to define conditioned reflex facts as those portions of conditioned reflex experiments which may be described in terms of stimulus substitution, and then states that conditioning theories are, or have been, based on such facts. Dashiell, on the other hand, accepts the Pavlov type of experiment as yielding conditioned response facts, regardless of whether these facts may be described in the limited manner indicated by Tolman. Certain data from conditioned response experiments, interpreted by Tolman as sign learning and therefore not conditioning, are accepted at their face value as conditioning facts by Dashiell because they arise in appropriate situations. While one of these interpretations is not necessarily more valid than the other, they differ in a manner which may cause confusion. It is important to be clear on two points: first, regarding what is meant by a conditioned response experiment, from which are derived 'conditioned response facts'; and, second, regarding the relation a conditioned response theory is to have to these facts.

The sources of conditioned response data will be limited for our purposes to laboratory experiments in which there occurs an alteration in response tendencies in respect to a (conditioned) stimulus by virtue of its repeated presentation in a controlled relationship with another (unconditioned) stimulus which with relative regularity evokes a response from the beginning of the experiment. The 'alteration in response tendencies' to be described as conditioning refers either to the appearance of a novel response or to the modification of a response originally evoked by the conditioned stimulus.[1] Whether the unconditioned response is to be described as a reflex is unimportant so long as the response consistently occurs. Delimiting the source of data to artificial laboratory situations has many theoretical advantages, and it need not prejudice a later consideration of the relationship between these facts and those discovered in familiar learning situations, such as memorization, acquisition of skill, the development of social attitudes. Furthermore, all relevant data recorded in such laboratory situations must be accounted conditioned response facts, whether or not they fit a conventional theory regarding the conditioned reflex.

It is a mistake to suppose that all theories built around the concept of conditioning attempt to solve the same problems, or apply to the same range of phenomena. Beginning with conditioning facts, as defined above, the elaboration of data may proceed along several lines. One may endeavor to discover the consistent relationships which hold within the data derived from a variety of conditioning experiments, and to formulate these relationships in intelligible expressions

[1] It is necessary to state the alteration as one in 'tendency' to respond, because such alterations are sometimes inferred rather than directly measured. The inferred alterations are not mythical or hypothetical; they are behavioral changes. For example, a given stimulus, following extinction, may evoke no response at all. Although this occurs in two situations, the two are not necessarily equivalent, for it is conceivable that in one case extinction proceeded well 'below zero.' The differences in response tendencies under the two conditions would have to be inferred from other events, *e.g.*, time required for recovery, inhibition-facilitation effects. Or overt performance may be the resultant of two tendencies, each of which is measurable by inference. To make such functional inferences is a very different thing from reifying the underlying physiological events.

which may be designated functional laws. This is the first stage in the elaboration of data. A few such relationships are already common knowledge: that by and large a forward order of conditioning is more favorable than a backward order, that non-reinforced conditioned responses tend to undergo extinction, and so on. These examples suggest the possibility of more precise formulation on the basis of experimental results, with little regard to theory.

This preliminary systematization, while not itself either based upon or producing a theory of conditioning, may prejudice the direction of later theorizing. The qualitative and quantitative rubrics selected for classification may greatly alter the findings from identical situations. There is no purely factual answer to the question: "When are two responses equivalent?" Responses are never exactly alike; by one criterion they may be equivalent, by another unequivalent. The adequacy of the descriptive categories chosen for purposes of systematization must be scrutinized with care if confusion is to be avoided in the further theoretical interpretation of factual data.

Granting this preliminary organization of data, two levels of theory may be indulged in. The first of these may be called theory of conditioning, the second, theory of learning based on conditioning.

1. *Theory of conditioning*—By a theory of conditioning is meant an attempt to explain the facts discovered in conditioning experiments. Such theories may be reductive or non-reductive.

Pavlov provides a theory of the interplay of excitatory and inhibitory processes in the cortex as the underlying causes of the phenomena revealed in the laboratory. Such attempts to understand conditioning in terms of neurophysiology are reductive, *i.e.*, the neurophysiological facts are considered basic to the interpretation of the behavioral facts.

Non-reductive theories seek to rewrite the specific equations of conditioning experiments in more general terms, so that a single equation may cover a greater variety of phenomena. Explanations are shorthand, clarified descriptions.

Conditioning is not reduced to neurophysiology, but is described by concepts appropriate to the actual experiments from which the data come. The scientific logic now commonly described as 'operational' is favorable to this type of theory.

The aims of the reductive and non-reductive theories are different; there need be no conflict between them. Neither of them, as thus described, is a theory of learning. They remain at the level of theories of conditioning.

2. *Theory of learning based on conditioning*—The behavioral facts of conditioning as systematized in functional laws may be used to deduce other types of behavior, many of which are superficially very unlike the conditioning experiment. This is the logical procedure by which Hull (**15**, and later) and Lepley (**17**) have been guided in utilizing conditioning principles in relation to learning. This is true also, but to a lesser extent, of Guthrie,[2] who is not interested in the physiology of conditioning, nor in the detailed relationships within experiments, but wishes primarily to interpret the applicability of conditioning principles to a wide range of phenomena.

It appears at once that an accurate delineation of theories of conditioning presupposes the results of systematization of data. That is to say, it is impossible to correlate conditioning with the facts of the nervous system until a precise statement of conditioning facts is at hand. To arrive at equations of higher generality, the equations expressing particular relationships must be dependable. To establish the relationship between conditioning facts and the phenomena of learning the data in both series must be in order. It is by no means clear that agreement has been reached regarding the functional laws holding within conditioning experiments; the relevant data may not even be at hand for stating some of the more important relationships. It is necessary to examine first the status of knowledge regarding conditioned responses before inquiring into the success of various theories falling within the above types.

[2] While Guthrie pays some attention to experiments on conditioning, his basic principles are not drawn from the Pavlov type of experiment, which he calls a special case of conditioning (**7**, 42 *ff*). This is a very important distinction between his explanatory procedure and that of Hull and Lepley.

Which relationships holding within experiments are to be considered essential, and which are to be regarded as accessory, depends in part upon the guiding principles used in systematization. Two such principles have been most widely used in developing conditioning theories or in refuting them. The first of these is the principle of *stimulus-substitution*, which Tolman rightly takes to be the orthodox expression of conditioning. The second principle may be called *organismic*, to use a term which places it in a systematic setting common to Gestalt theory and to some of the contemporary movements in biology (3). Tolman discovers in conditioning experiments many instances of sign-learning, which he believes cannot properly be comprehended in terms of the conventional stimulus-substitution analysis, but require Gestalt-like laws. More broadly conceived, the organismic, sign-learning, or Gestalt interpretations of conditioning are essentially teleological, while the stimulus-substitution concept is in essence mechanistic. It is this latter concept which will be treated in this paper.

The Case for Stimulus-Substitution

The diagram of Fig. 1 is a familiar representation of the conditioning process. It shows that the previously neutral conditioned stimulus acquires the properties of the unconditioned stimulus. The dog which once salivated only to food learns to salivate also to a bell. To the extent that the conditioned stimulus replaces the unconditioned one, the diagram illustrates what may be designated the *stimulus-substitution formula.*

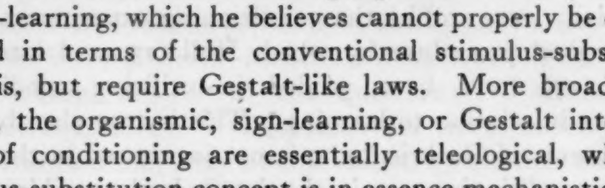


Fig. 1. Conventional substitute-stimulus diagram.

Since the conditioned stimulus is never in fact strictly neutral, the diagram of Fig. 1 is often elaborated one step by adding an original response to the conditioned stimulus, as in

Fig. 2.[3] This diagram, somewhat more realistic than the first, already clouds the mechanical simplicity of the conditioning process by posing the question as to why one diagonal is favored over another. It is necessary to adduce supplementary principles (intensity? time order? dominance?) to account for the direction of conditioning. These considerations apart, both diagrams agree in the fundamental conception that the conditioned stimulus acquires the potentiality of evoking the response originally elicited by the unconditioned stimulus.

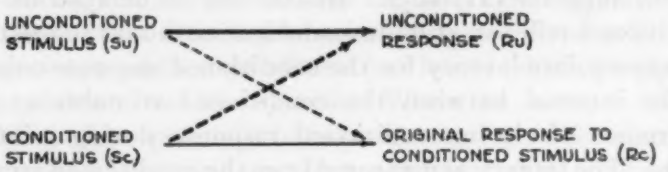


FIG. 2. Substitute-stimulus diagram modified to include original responses to the conditioned stimulus.

These diagrams seem to indicate that the conditioned responses are in every respect similar to the unconditioned responses. Watson early stressed the point that "the general and coarser motor features are closely similar" (28). Lashley implied that the completely integrated response was shunted to the new stimulus (16). Critics of the stimulus-substitution formula (9), (14), (18), (23), (29) have used in evidence against these assertions the fact that conditioned responses differ in latency and amplitude from the unconditioned responses. Such evidence alone does not suffice as a contradiction of stimulus-substitution, for even the most hide-bound defender of this view would admit that conditioned responses as recorded in the laboratory might be expected to differ from the unconditioned responses in certain important respects. Among these expected points of difference would be amplitude, latency, and configuration, the very aspects which critics have declared to be damaging to the stimulus-substitution formula. The argument that differences in these respects are not

[3] These diagrams correspond to the Type II conditioned reflex described by Skinner (24). His Type I, not specifically represented in Pavlov's work, is omitted from consideration at this point for purposes of simplicity, although the definition of conditioning adopted permits its inclusion.

damaging to the principle of stimulus-substitution is somewhat as follows:

1. *Latent time*—The latency of the conditioned response cannot be expected to remain the same as that of the unconditioned response, even though established by simultaneous conditioning, because the conditioned stimulus may activate sensory systems with temporal characteristics different from those activated by the unconditioned stimulus. The central mechanisms are undoubtedly more complex, as Pavlov himself suggests (21, 105). In the case of delayed or trace conditioned reflexes, stimulus-substitution would suggest that the appropriate latency for the conditioned response ought to be the interval between the conditioned stimulus and the occurrence of the unconditioned response during reinforcement. The latency as measured from the conditioned stimulus would obviously be longer than that measured from the unconditioned stimulus.

2. *Amplitude*—Because the conditioned stimulus usually differs in intensity from the unconditioned stimulus, and affects a sense organ with different characteristics from that affected by the unconditioned stimulus, one may expect a different effective intensity for the two stimuli. The lesser effectiveness of the conditioned stimulus may be due either to its physical characteristics in relation to the sensory system, or to the less well-established relationship existing between it and the response in the process of acquisition. In either case the conditioned response may be expected to be of less amplitude than the unconditioned response, without doing violence to the substitution principle.

3. *Configuration*—There are at least two reasons why the conditioned response may be expected to differ in form from the unconditioned response. In the first place it is well known that differences in intensity of stimulation produce responses not only of greater amplitude but also of altered pattern, including such phenomena as after-discharge. To the extent that these characteristics are related to intensity, configurational differences follow as a corollary to the admission of intensity differences. In the second place, certain aspects of the

unconditioned response depend upon the manipulation of the unconditioned stimulus, and cannot be given to the conditioned stimulus in isolation.[4] The dog is unable to chew and swallow following the ringing of the bell when food is withheld in exactly the manner possible with food present. Stimulus-substitution can demand the elicitation by the conditioned stimulus of only those portions of the unconditioned response which are free to be given in the absence of the unconditioned stimulus.

Guthrie, who has given one of the most careful elaborations of a theory of conditioning based on the principle of substitute stimulation, could accept all that has been said above without doing violence to his system. While he accepts as his fundamental proposition that the cues acting at the time of a response tend on their occurrence to elicit that response (7, 26), he proceeds to show that the conditioned response will not be just like the unconditioned response. When conditioned and unconditioned stimuli occur together, there results a response which is a compromise between the movements evoked by the conditioned and unconditioned stimuli. What gets conditioned is the movement which actually occurs. Because of the compromise character of the responses occurring during reinforcement, the conditioned response may be expected to differ from the unconditioned response. It is also different because of the necessarily abortive nature of the responses in the absence of the unconditioned stimulus (7, 79). Thus Guthrie's analysis agrees with ours in that a demonstration of difference in descriptive character between conditioned and unconditioned responses is insufficient evidence against stimulus-substitution.

Critics of the stimulus-substitution formula have also emphasized the appearance of the conditioned response during reinforcement as a new response. They say that it is neither the old unconditioned response, nor the compromise movements present when both conditioned and unconditioned stimuli occur together at the beginning of the experiment

[4] This manipulatory intermediary is the fundamental principle involved in Skinner's Type I conditioned reflex. In certain respects the Types I and II approximate each other, as Skinner has pointed out (24).

(9), (11). When the time relations between conditioned and unconditioned stimuli permit, this new response often occurs just before the unconditioned stimulus. For this reason it is sometimes referred to as an anticipatory reaction.[5]

The diagrams presented in Figs. 1 and 2 do not represent what actually occurs in most conditioning experiments. They fail to show that simultaneity is a less favorable order of presentation of stimuli than succession; they fail to show that conditioned and unconditioned responses may occur separately during reinforcement, the conditioned response often preceding the unconditioned response. A better summary of the facts is given in Fig. 3. This is not a diagram of events in

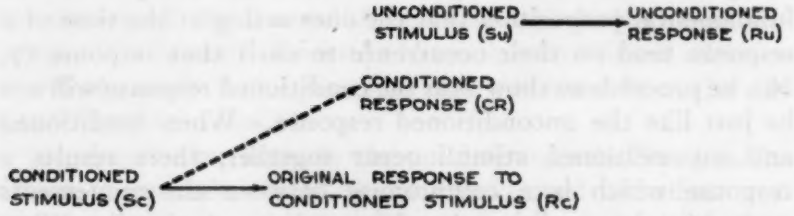


FIG. 3. Diagram of conditioning modified to indicate sequence of events commonly found in experiments. The problem is whether or not the conditioned response (CR) is equivalent to the unconditioned response (R_u).

the nervous system, and it does not present a theory of conditioning. It merely outlines what happens frequently in experiments on salivation, galvanic skin responses, motor alimentation, breathing, eyelid reactions, knee-jerks, finger withdrawal, pupillary responses. The conditioned response often occurs as the first component of a dual response, the second component of which is the unconditioned reaction, as in the second and third records of Fig. 4. The question at issue is whether the intrusion of the conditioned response as a new reaction, often anticipatory, contradicts the predictions based on the formula of stimulus substitution.

[5] The term is unfortunate because it has two meanings. To anticipate means to precede; it also means to expect. Only the former meaning is implied here, as in the writings of Hull (15, and later). Whether or not the second meaning, of expectation, is also applicable, will be considered more fully in a later paper dealing with teleological aspects of conditioning.

An experimental situation will help to focus the problem. Suppose that a buzzer has sounded regularly half a second before a tap on the patellar tendon elicits a knee-jerk, until there has developed a conditioned tendency to kick the leg to the buzzer. If the tap is omitted, when, according to the stimulus-substitution principle, should the leg movement occur? One hypothesis would be that the situation is redintegrated in as near its original form as possible; that is to say, the kick follows the buzzer when the tap is omitted by the same interval it usually followed the buzzer, slightly more than half a second. The occurrence of the anticipatory reaction makes this hypothesis invalid. This is not, however, the only hypothesis consistent with the notion of stimulus substitution. Another hypothesis (possibly demanding less from the organism than that above) is that once a conditioned tendency is developed, the time for the conditioned response to occur is shortly after the onset of the conditioned stimulus. That is to say, the knee-jerk will tend to occur at approximately reaction-time latency following the buzzer. A more precise temporal adjustment might wait upon longer continued training. If the conditioned stimulus is thought of as a sort of trigger mechanism to set off the response, the fact that the leg might be in motion following the buzzer before the tap is delivered to the tendon causes no embarrassment for the substitution theory. One might expect considerable fluctuation in the latency of the conditioned responses during the process of substitution, before stabilization occurred. Many of these responses would tend to happen within half a second of the buzzer; these would be anticipatory. There is experimental evidence that such a process takes place, as in Switzer's study of galvanic responses.

In Switzer's experiments a light flashed on 16 seconds before a shock evoked the unconditioned galvanic skin response. The first conditioned responses occurred not 16 seconds after the light, but very promptly, at latencies of the order of 5 seconds (**25**, 613–615). Only after continued reinforcement at the 16-second interval was a delay established, so that some 11 seconds elapsed after the light before

the conditioned response occurred. Switzer's results, which show extreme early anticipation, could be predicted in terms of the second hypothesis offered above: the light, having newly acquired the capacity to evoke a response, elicits it promptly. Delay is a refinement which requires time.[6]

Another series of experiments favoring the hypothesis that substituted responses may occur at low latency, and thus appear anticipatory, are those of Culler and his associates (5). They report as the first thing to happen in their efforts to secure from dogs conditioned leg withdrawals to tone something best described as stimulus substitution. When the tone is turned on the dog behaves as though he were being shocked. These are 'anticipatory reactions.' Later in the experiments there develops a different kind of reaction, which Culler and his co-workers call 'adaptive,' displacing the earlier form. The point to be noted here is that the anticipatory reaction instead of being contradictory, more nearly agrees with the expectations of stimulus-substitution.

Thus far it has been shown that differences in descriptive characteristics between responses are not damaging to the principle of stimulus substitution, and that the fact of anticipation cannot of itself be construed as a violation of the formula. Other evidence must be marshalled if a case is to be made against the substitute-stimulus paradigm.

[6] These experiments of Switzer's bear importantly on Guthrie's theory regarding delayed conditioned responses. In order to preserve simultaneity of cue and response as the only basis for conditioning, Guthrie posits the belief that the true conditioner in a situation involving delay is not the designated conditioned stimulus, but rather the proprioceptive consequences of reactions set off by this stimulus. These proprioceptive cues, occurring along with the unconditioned response, are the true conditioned stimuli (7, 53 *ff*). Switzer's results deny Guthrie's fundamental proposition that simultaneity of cues and response is essential, because it is evident that the low-latency conditioned response in the early stages of the experiment cannot be set off by cues present only 16 seconds after the light flashes on. On the other hand, the delay which gradually develops may well be implemented with cues such as Guthrie suggests. The approximate simultaneity of cues and response, which Guthrie would have primary, turns out to be a secondary elaboration of the conditioning process.

Switzer's finding that even after the delay is established the response continues to anticipate the unconditioned stimulus is completely in harmony with Pavlov's work.

Against Stimulus-Substitution

It is implicit in the above discussion that the anticipatory reaction is the unconditioned response attached to a new stimulus, displaced in time, and modified in certain quantitative ways which are not essential to its identification as the old response. Both conditioned and unconditioned responses may occur during reinforcement because two stimuli occur in succession, each of which is capable of eliciting the response. The presence of both reactions is no more remarkable than two knee-jerks to two successive blows on the tendon. Pertinent criticism of stimulus-substitution must discover differences between the conditioned and unconditioned responses which are more significant than their similarities. These differences may be either in regard to descriptive characters, or in regard to functional relationships. That such differences do in fact exist is the thesis now to be defended.

1. *Descriptive differences between conditioned and unconditioned responses*—It has already been pointed out that differences in latency, amplitude, or configurational characteristics, although conclusively demonstrated, are not convincing evidence that conditioned and unconditioned responses are essentially unlike. The argument with regard to configuration was limited, however, to those differences in configuration which are correlated either with intensity and amplitude, or with mechanical possibility in the absence of the unconditioned stimulus. There are configurational differences discovered in experiments which are beyond the limits set by these arguments. Among responses showing such differences are those which are in some sense the opposite of the unconditioned responses. Conditioned breathing is illustrative; breathing following the onset of the conditioned stimulus is restricted, whereas following the unconditioned stimulus it is enhanced. The same is sometimes true of conditioned galvanic skin responses; the resistance change may be in opposite directions following conditioned and unconditioned stimuli.

It is important to realize that the restrictions imposed by experimental circumstances tend to make the conditioned response seem more like the unconditioned response than it

may be in fact. When a single response is recorded the only changes which are measured are changes in that response; occasionally changes in only one direction are indicated. If finger withdrawal is arranged to break an electrical circuit, depression of the finger following the conditioned stimulus will not be recorded, and the only movements registered will be those like the unconditioned withdrawal response. Even where either increase or decrease in magnitude of response is registered, similarity in these changes may be accompanied by other response changes which will deny the organic similarity between conditioned and unconditioned responses. Successful conditioning is frequently considered to have been obtained only if the conditioned response is similar in gross characteristics to the unconditioned response; some subjects, who do not yield such responses, do not 'condition.' Undoubtedly many of the subjects who do not 'condition' in the finger withdrawal experiment, in the very same experiment develop conditioned breathing reactions. The latter may or may not have a chance to appear as data. For the purposes of the particular experiments, it is methodologically permissible to record any aspects of the situation in which the experimenter happens to be interested; in an effort to understand what actually takes place, it is often necessary to know more than the records show. With full realization of these restrictions imposed by laboratory methods, those relatively rare experiments in which the conditioned response moves off in the opposite direction from the unconditioned response are of more importance than their frequency implies.

The difference between the anticipatory response and the unconditioned reinforcing reaction is emphasized in the experiments of Upton (27) and Britt (4) on conditioned breathing in the guinea-pig and rat, respectively. Both find two types of conditioned response. The more prevalent response is reduced breathing although an occasional one is more nearly a replica of the exaggerated breathing which normally follows shock. The 'true' conditioned responses appear only late in the conditioning, are very rare and are not typical. Horton (13), working with a technique similar to Upton's, with the

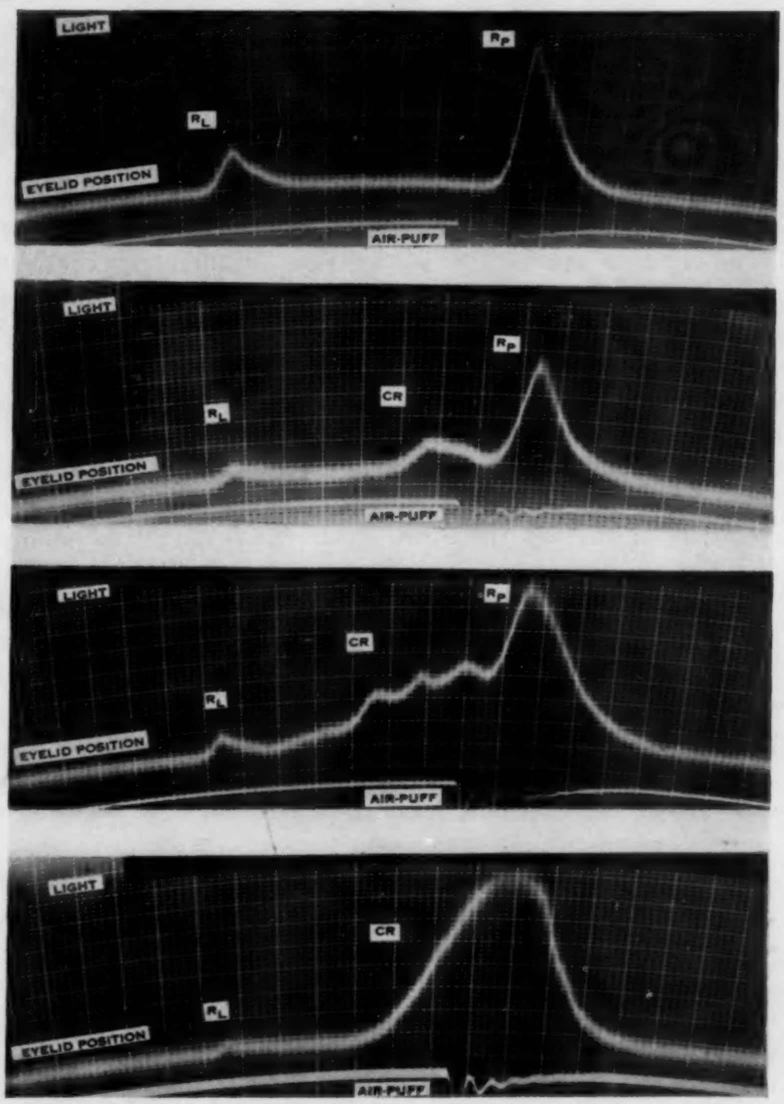


FIG. 4. Variations in conditioned eyelid responses of one subject. The upper record illustrates the reflexes to light (R_L) and to the air-puff (R_P) before there is any evidence of conditioning. The conditioned response (CR) appears at first as a reaction of small amplitude, as in the second record. The third and fourth records show variations in the conditioned response as it becomes better established. The recruitment type of response (third record from the top) was the more frequent for this subject. All records are from one human subject, secured in an investigation by Hilgard and Campbell (**10**). Vertical time lines, 10 ms., emphasized lines, 50 ms. The air-puff follows the light by approximately 400 ms.

same species, failed to secure them at all. Regardless of the disposition to be made of the 'true' conditioned responses (*i.e.*, those following the predictions of stimulus-substitution), these authors indicate clearly that the more usual conditioned response is quite unlike the unconditioned reflex.

In Fig. 4 there are presented specimen records from a human subject showing conditioned eyelid reactions moving in the same direction (closing) as the conditioned response, but unlike it in ways not to be accounted for either by intensity of stimulation, or by the impossibility of duplicating the unconditioned response in the absence of the normal stimulus. The top record shows the responses before conditioning. In the second record there appears the 'anticipatory' type of conditioned response, which is possibly conceived of as a substitute for the unconditioned reflex to the air-puff. In the third record from the top the form of the conditioned response is distinctly different from that of the unconditioned reflex. The response shows a gradual recruitment never found in the reflex to the air-puff. The record at the bottom shows a type of response given occasionally by this subject late in the conditioning series. It is a single smooth response, beginning before the unconditioned stimulus and incorporating the unconditioned reflex. It is evident that more needs to be said about these conditioned responses than is implied in stimulus-substitution.

2. *Functional relationships indicating differences between conditioned and unconditioned responses*—In its most dogmatic expression, the formula of stimulus-substitution demands that an anticipatory conditioned response and a following unconditioned response are successive reactions of essentially the same kind. Such differences as exist are trivial and nonessential. If, however, the differences indicate that the sequence of events is indeed an organized series, something must be added to the substitution formula in order to express the dynamics of the process.

When similar reflexes occur in succession to adequate stimuli spaced at intervals favorable to conditioning, the second of the two reflexes is greatly depressed in magnitude,

even though the first of the two reflexes is of inconsiderable amplitude. That a full-fledged conditioned response may intrude between these two reflexes, and even summate with the second of them, has been demonstrated by experiment (**9**). The records of Fig. 4 are illustrative. This is behavioral evidence that the conditioned response is functionally something other than the first of two similar reflexes.[7]

That the forward order of stimuli is more favorable than absolute simultaneity is another evidence for some kind of organization inherent in the conditioning process, of which the anticipatory response is but one evidence. The anticipatory response might be an artifact due to the chosen interval, but it is significant that the most favorable intervals for conditioning have been found to be those in which the margin between conditioned and unconditioned stimuli is sufficient for the anticipatory response to occur. Shorter intervals, even in the forward order, are as unfavorable as backward conditioning (**2**), (**31**). Succession of stimuli rather than coincidence is probably fundamental in the dynamics of conditioning. Of itself the fact of succession says nothing about a difference between conditioned and unconditioned responses, but it becomes pertinent when linked to the facts of experimental extinction.

Experimental extinction is not only one of the best attested principles of conditioning,[8] but it is one of the most illuminating for revealing the nature of the process. The maintenance of the conditioned response is contingent upon reinforcement; non-reinforced repetition results in decreased conditioned responses. These facts require some sort of dynamic or functional supplementation of the stimulus-substitution concept. If the new stimulus simply substituted

[7] This is not to say that the conditioned response exerts no inhibitory influence on the unconditioned response. To unravel the exact relationships would require an extensive search into the physiology of facilitation and inhibition. It may be remarked that the interaction is more like that between reflex and voluntary reactions than that between two similar reflexes. See, for example, Peak (**22**).

[8] It is only fair to note that failure to secure extinction has been reported in several experiments (**8**), (**12**), (**14**), (**20**). These failures are not to be lightly dismissed, but they have not yet been sufficiently analyzed. Two of the cases (**14**), (**20**) involve conditioning in which words are involved as stimuli.

for the old, further reinforcement would seem to be unnecessary. The responses which are conditioned, according to Guthrie, are the responses which occur. Once the conditioned responses have occurred, they ought to be self-perpetuating by this principle. That they do not perpetuate themselves is recognized, and additional hypotheses have been offered. Two such hypotheses, by Winsor (**30**) and by Guthrie (**7**), may be mentioned.

Winsor believes that the omission of the unconditioned stimulus results in responses which have not occurred during the reinforcement trials. These new responses act as distractors or inhibitors under Pavlov's principle of external inhibition, and in subsequent trials the responses to the conditioned stimulus are decreased. This plausible theory assumes, however, that there is some function served by the unconditioned response which cannot be served by the conditioned response, so that the omission of the unconditioned stimulus is disturbing. This is already a dynamic supplementation of mere substitute-stimulation.

Guthrie introduces a further emendation to account for extinction. When reinforcement does not occur the attitude, posture, or 'set' of the animal is altered. Spontaneous recovery is possible because with lapse of time the original 'set' is reinstated. This notion of attitude or set adds an organizing principle to the process of extinction which logically must also be important in the formation of conditioned responses. In this organization under attitude or set, conditioned and unconditioned stimuli and responses play unlike roles. To the extent that the responses are functionally unlike, they differ from each other, no matter how similar they may be in descriptive characteristics.

The theories of both Winsor and Guthrie agree in a general way with the facts of conditioning, although both theories require confirmation in detail. Neither seems quite adequately to account for the progressive nature of the changes during extinction. Furthermore, the demonstration of external inhibition, required by Winsor's theory, is very difficult even in responses which extinguish readily (**19**). It suffices

for the present, however, to accept the theories as plausible, but to point out that they describe the conditioning process in terms other than simple stimulus-substitution.

It is pertinent at this point to note an observation of Bekhterev's that a dog whose foot was tied down so that he could not respond in the usual fashion to the conditioned stimulus, responded instead with the other foot (1, 216). This suggestion of adaptive significance of the events goes beyond stimulus-substitution, and leads to the problems of the functional organization of the conditioning process, reserved for consideration in another paper.

Summary

Conditioned response facts derive from conditioning experiments. A conditioning experiment is one in which there occurs an alteration in response tendencies in respect to a (conditioned) stimulus by virtue of its repeated presentation in a controlled relationship with another (unconditioned) stimulus which with relative regularity evokes a response from the beginning of the experiment.

Theories revolving about conditioning fall into two types: (1) those which attempt to explain conditioning facts, and (2) those which attempt to deduce other types of behavior, such as familiar forms of learning, from the facts of conditioning. The first type may be called theories of conditioning, the second, theories of learning (or of other behavior) based on conditioning.

Systematization of the facts within experiments presupposes a logical procedure. Some writers have utilized as a guiding principle the notion of stimulus-substitution, while others, especially critics, have attempted to view the facts organismically. This paper is concerned with the attempt to organize the facts around the stimulus-substitution formula.

The substitute-stimulus conception of conditioning is given support by many of the descriptive facts of conditioning, some of which have been supposed to contradict it. On the other hand, conditioned responses possessing a configuration differing widely from the unconditioned responses on which

they are based are not predictable in terms of stimulus-substitution. The dynamics of the conditioning process require either the abandonment of the simple formula, or its supplementation with other concepts which will account for descriptive differences in the responses; for the fact that succession of stimuli is more favorable than simultaneity; and for the tendency towards extinction of conditioned responses when not reinforced.

REFERENCES

1. Bekhterev, V. M., General principles of human reflexology (Trans. by E. and W. Murphy), New York: International Publishers, 1932, pp. 467.
2. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. Gen. Psychol.*, 1934, **10**, 173–197.
3. v. Bertalanffy, L., Modern theories of development (Trans. and adapted by J. H. Woodger), London: Oxford University Press, 1933, pp. x + 204.
4. Britt, S. H., Tonal sensitivity in the white rat, *J. Comp. Psychol.*, 1935, **19**, 243–264.
5. Culler, E., Finch, G., Girden, E. and Brogden, W., Measurements of acuity by the conditioned response technique, *J. Gen. Psychol.*, 1935, **12**, 223–227.
6. Dashiell, J. F., A survey and synthesis of learning theories, *Psychol. Bull.*, 1935, **32**, 261–275.
7. Guthrie, E. R., The psychology of learning, New York: Harper and Bros., 1935, pp. viii + 258.
8. Harlow, H. F. and Stagner, R., Effect of complete striate muscle paralysis upon the learning process, *J. Exper. Psychol.*, 1933, **16**, 283–294.
9. Hilgard, E. R., Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound, *Psychol. Monog.*, 1931, **41**, No. 184, pp. 50.
10. —— and Campbell, A. A., The course of acquisition and retention of conditioned eyelid responses in man, *J. Exper. Psychol.*, 1936, **19**, 227–247.
11. —— and Marquis, D. G., Acquisition, extinction, and retention of conditioned lid responses to light in dogs, *J. Comp. Psychol.*, 1935, **19**, 29–58.
12. ————, Conditioned eyelid reactions in monkeys, with a comparison of dog, monkey, and man, *Psychol. Monog.*, 1936 (to appear).
13. Horton, G. P., A quantitative study of hearing in the guinea pig (*Cavia cobaya*), *J. Comp. Psychol.*, 1933, **8**, 3–51.
14. Hudgins, C. V., Conditioning and the voluntary control of the pupillary light reflex, *J. Gen. Psychol.*, 1933, **8**, 3–51.
15. Hull, C. L., A functional interpretation of the conditioned reflex, Psychol. Rev., 1929, **36**, 498–511.
16. Lashley, K. S., Studies of cerebral function in learning. VI. The theory that synaptic resistance is reduced by the passage of the nervous impulse, Psychol. Rev., 1924, **31**, 369–375.
17. Lepley, W. M., A theory of serial learning and forgetting based on conditioned reflex principles, Psychol. Rev., 1932, **39**, 279–288.
18. Liddell, H. S., The conditioned reflex, 'Comparative psychology' (edited by F. A. Moss), New York: Prentice-Hall, 1934, 247–296.

19. Miller, A. R., A failure to confirm Pavlov's hypothesis of external inhibition, *Amer. J. Physiol.*, 1934, **108**, 608–612.
20. Osipova, V. N. (Unextinguished associated reflexes in children) (Russian). Reference from Razran, G. H. S., Conditioned responses in children, *Arch. Psychol.*, 1933, **23**, No. 148, p. 112.
21. Pavlov, I. P., Conditioned reflexes (Trans. by G. V. Anrep), London: Oxford University Press, 1927, pp. xv + 430.
22. Peak, H., Reflex and voluntary reactions of the eyelid, *J. Gen. Psychol.*, 1933, **8**, 130–156.
23. Schlosberg, H., Conditioned responses in the white rat, *J. Genet. Psychol.*, 1934, **45**, 303–335.
24. Skinner, B. F., Two types of conditioned reflex and a pseudo-type, *J. Gen. Psychol.*, 1935, **12**, 66–77.
25. Switzer, S. A., Anticipatory and inhibitory characteristics of delayed conditioned reactions, *J. Exper. Psychol.*, 1934, **17**, 603–620.
26. Tolman, E. C., Theories of learning, 'Comparative psychology' (edited by F. A. Moss), New York: Prentice-Hall, 1934, 367–408.
27. Upton, M., The auditory sensitivity of guinea pigs, *Amer. J. Psychol.*, 1929, **41**, 412–421.
28. Watson, J. B., The place of the conditioned reflex in psychology, Psychol. Rev., 1916, **23**, 89–116.
29. Wendt, G. R., An analytical study of the conditioned knee-jerk, *Arch. Psychol.*, 1930, **19**, No. 123, pp. 97.
30. Winsor, A. L., Observations on the nature and mechanism of secretory inhibition, Psychol. Rev., 1930, **37**, 399–411.
31. Wolfle, H. M., Conditioning as a function of the interval between the conditioned and the original stimulus, *J. Gen. Psychol.*, 1932, **7**, 80–103.

[MS. received November 11, 1935]

PSYCHOLOGICAL REVIEW PUBLICATIONS

Original contributions and discussions intended for the Psychological Review should be addressed to

Professor Herbert S. Langfeld, Editor PSYCHOLOGICAL REVIEW,
Princeton University, Princeton, N. J.

Original contributions and discussions intended for the Journal of Experimental Psychology should be addressed to

Professor Samuel W. Fernberger, Editor JOURNAL OF EXPERIMENTAL PSYCHOLOGY,
University of Pennsylvania, Philadelphia, Pa.

Contributions intended for the Psychological Monographs should be addressed to

Professor John F. Dashiell, Editor PSYCHOLOGICAL MONOGRAPHS,
University of North Carolina, Chapel Hill, N. C.

Reviews of books and articles intended for the Psychological Bulletin, announcements and notes of current interest, and *books offered for review* should be sent to

Professor John A. McGeoch, Editor PSYCHOLOGICAL BULLETIN,
Wesleyan University, Middletown, Conn.

Titles and reprints intended for the Psychological Index should be sent to

Professor Walter S. Hunter, Editor PSYCHOLOGICAL INDEX,
Clark University, Worcester, Mass.

All business communications should be addressed to

Psychological Review Company, Princeton, New Jersey

AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by K. M. Dallenbach, M. F. Washburn, Madison Bentley, and E. G. Boring.
Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Child behavior, animal behavior, comparative psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Herbert S. Langfeld.
Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per vol. 500 pages. Edited by John F. Dashiell.
Without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 400–500 pages. Edited by Walter S. Hunter and R. R. Willoughby.
An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by John A. McGeoch.
Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth.
Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 pages annually. Edited by Henry T. Moore.
Quarterly. Abnormal and social. Founded 1906.

Psychological Clinic—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witmer.
Without fixed dates (Quarterly). Orthogenics, psychology, hygiene. Founded 1907.

Journal of Educational Psychology—Baltimore: Warwick & York.
Subscription $6.00. 720 pages. Monthly except June to August.
Edited by J. W. Dunlap, P. M. Symonds and H. E. Jones. Founded 1910.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe.
Quarterly. Psychoanalysis. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J.; Psychological Review Company.
Subscription $7.00. 900 pages annually. Edited by Samuel W. Fernberger.
Bi-monthly. Experimental psychology. Founded 1916.

Journal of Applied Psychology—Indianapolis; C. E. Pauley & Co.
Subscription $6.00. 600 pages annually. Edited by James P. Porter, Ohio University, Athens, Ohio. Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Ed. by Knight Dunlap and Robert M. Yerkes. Two volumes a year. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Roy M. Dorcus, Editor.
Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
Subscription $7.00 per vol. One volume per year. Edited by Carl Murchison.
Bi-monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages ann. Edited by Walter S. Hunter and R. R. Willoughby.
Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Experimental, theoretical, clinical, historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Ed. by John Dewey and Carl Murchison.
Quarterly. Political, racial, and differential psychology. Founded 1929.